


```

34  rglleProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
104  GGATATCCTGTCCCCACGCTGTGGAGAGCCGCCAGCACTAACCAGGTT 153
51  ValGlyGlyGluAspSerThrAspSerGluTrpProTrrIleValSerI 67
154  GTGGCGGGGAGGAGACGACACGAGGAGGAGGAGGAGGAGGAGGAGGAG 203
67  eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArgT 84
204  CCAGAGAGATGGAGACCCACGCTGCCAGGTTCTCTCTCAACGAGCGCT 253
84  rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
254  GGTGATCATCTGCTGCCACGCTGTTCAAGGACACGAGCAACCAACATAC 303
101  LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyLysProGlySerAr 117
304  CTGTCTCTGTGCTGTGGGGGCTGAGCTGGGGAACCTGCTCTCTG 353
17  gSerGlnLysValGlyAlaAlaTrpValGluProHisProValTyrSerT 134
354  GTCCAGAGAGTGGGTGTGTGCTGTGGAGAGCCGCCAGCTGTATTCCT 403
134  rPlyGlyGlyAlaCysAlaAspIleAlaLeuValAlaGluGluArgSer 150
404  GGAAGGAGAGGCTGTGCAGACATGCTGCTGCTGCTGCTGCTGCTGCTG 453
151  IleGlnPheSerGluArgValAlaLeuProIleCysLeuProAspAlaSerI 167
454  ATACAGTTCACAGAGCGGCTCTGCCATCTGCTGCTGCTGCTGCTGCTAT 503
167  eHisLeuProProAsnThrHisCysTrpIleSerGlyTrrIleGlySerI 184
504  CCACCTCTCTCAAAACACCCAGCTGATCTCAGGCTGGGGAGAGATCC 553
184  LAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
554  AAGTGGAGTTCCTTGGCCGCCACCTGACCTGACAGGAGGAGGAGGAG 603
201  ProIleIleAspSerGluValLysSerHisLeuTrrIleArgIleVal 217
604  CCAATCATCTGCTGAGAGCTGCTGACGCACTCTACTGCGGGGAGCAGG 653
217  YGlnGlyProIleThrGluAspMetLeuCysAlaGlyTrrLeuGluGly 234
654  ACAGGAGACCATCATGAGGACATGCTGTGCTGCTGCTGCTGCTGCTG 703
34  LuArgAspAlaCysLeuGlyAspSerGlyProLeuMetCysGlnVal 250
704  AGCGGAGTCTGTCTGGGGAGCTCGGGGGGCCCTCTCATGTGCCAGGTG 753
251  AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlyCysAl 267
754  GAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 803
267  aGluArgAsnArgProGlyValTrrIleSerLeuSerAlaHisArgSerT 284
804  CGAGCGCAACAGGCCCGGGGCTTACATCAGCTCTGCTGCGCACCGCTCT 853
284  rPValGluLysIleValGlnGlyAlaGlnLeuArgGlyValArgIleGly 300
854  GGGTGGAGAGATCGTGCAGAGGGGTGCACTCCGGGGGGGCGCTAGGGG 903
301  GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
904  GGTGGGGGCGCTCAGGGGACCGAGGAGGAGGAGGAGGAGGAGGAGGAG 953
317  r 317
954  C 954

```

seq_name: gb:AF321182

seq_documentation_block:

LOCUS AF321182 1332 bp mRNA linear PRI 26-DEC-2001

DEFINITION Homo sapiens serine protease PRSS22 mRNA, complete cds.

ACCESSION AF321182

VERSION AF321182.1 GI:11386012

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1332)

Wong G.W., Yasuda S., Madhusudan M.S., Li L., Yang Y.,

Krill S.A., Salih A. and Stevens R.L.

Human trypsin epsilon (PRSS22), a New Member of the Chromosome

16p13.3 Family of Human Serine Proteases Expressed in Airway

Epithelial Cells

JOURNAL J. Biol. Chem. 276 (52), 49169-49182 (2001)

REFERENCE 2 (bases 1 to 1332)

Wong G.W.

Direct Submission

Submitted (14-NOV-2000) Rheumatology, Immunology and Allergy,

Brigham and Women's Hospital, Harvard Medical School, 1 Jimmy Fund

May, Boston, MA 02115, USA

FEATURES

source

1.1332

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16p13.3"

/tissue="pancreas"

18.971

/codon_start=1

/product="serine protease PRSS22"

/protein_id="AF35070.1"

/db_xref="GI:11386013"

/translation="MYSGAPALGSGGCTGTTSTLLASTAIIAARIPPPACGK

COILNRVGGEDSDSDSEWPVIVSIQKNGTHKAGSLITSRWVITTAHCRDNKPYLF

STLGMQAGNRSKQKGVAMVEHPVYSKEACADIALVRLERSIQSERVLP

CLPDASIHLPNTHWISGWSIQDVPYLPHPOTLOKLPIDSEVSHLYMRGAG

GPTEMLACAGLYEGRDACLDSGGGLMCOYDGLMGLAGLISMEGCAERNRPVYI

SLSAHSRWEKIVOGVQLRRAQGGALAPSGGGAARS"

BASE COUNT 223 a 445 c 403 g 261 t

ORIGIN

alignment_scores:

Quality: 1720.00 Length: 317

Ratio: 5.426 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-040-803-7 x AF321182 ..

Align seg 1/1 to: AF321182 from: 1 to: 1332

1 MetValValSerGlyAlaProProAlaLeuGlyGlyGlyCysLeuGlyTh 17

18 ATGTGTGTTTGGAGAGCGCCGCCAGCCCTGGGGGGGCTGTCTGGCAC 67

17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34

68 CTTCACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 117

34 rglleProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50

118 GGATACCTGTTCCGCCAGCTGTGGAGAGCCCGACGAGAGCTGACCG 167

51 ValGlyGlyGluAspSerThrAspSerGluTrpProTrrIleValSerI 67

168 GTGGCGGGGAGGAGACGACACTGACAGGAGTGGCCCTGATCTGTAGCAT 217

```

67  eglLysasnGlyThrHisCysAlaGlySerLeuLeuThrSerArgT 84
218  CCAGAGAAATGGACCACTGGCGAGGTTCTCTCTACCAAGCGGCT 267
84  rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
268  GGGGATGACTGTGGCTGCTGCTTCAAGGACAACCTGAAACAAACATAC 317
101  LeuPheSerValLeuLeuGlyAlaTTrpGlnLeuGlyLysnProGlySerAr 117
318  CTGTCTCTGTGCTGCTGGGGCTGAGTGGGGAACCTGGCTCTCG 367
117  gSerGlnLysValGlyAlaTTrpValGluProHisProValTyrSerT 134
368  GTCCCAAGAGTGGGTGTGCTGGTGGGAGGCCACCTGTGTATTCCT 417
134  rPlySGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
118  GGAAGGAAGTGGCTGTGACAGACATGCCCCGTGCGTGCAGACGCTCC 467
151  IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167
468  ATACAGTCTCAGACGGGCTCTCCATCTGCTACCTGATGCTCTAT 517
167  eHisLeuProAsnThrHisCysTrpIleSerGlyTTrpGlySerIleG 184
518  CCACCTCCCTCCAAACACCCACTGTGTATTCAGCTGGGGGAGCATCC 567
184  InasPglYValProLeuProHisProGlnThrLeuGlnLysLeuLysVal 200
568  AAGATGAGTTCCTCCCTGGCCACCTCAGACCTGACAGAACCTGAAGT 617
201  ProIleIleAspSerGluValCysSerHisLeuTyrTrpArgValAlaG 217
618  CCTATCATCATGACTGGAAAGTCTGCAAGCCATCTGTACTGCGGGAGG 667
217  gLingIlyProIleThrGluAspMetLeuCysAlaGlyTTrpLeuGluG 234
668  ACAGGAGCATCATCTGAGACATGCTGTGTGCGGCTACTTGGAGGGG 717
234  LuArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnVal 250
718  AGCGGAGATGCTGTGCTGGGCGACTCGGGGGCCCCCTCATCTATGCCAG 767
251  AspGlyAlaTTrpLeuLeuAlaGlyTTrpIleSerTrpGlyGluGlyC 267
68  GACGGCGCTGCTGCTGGCGCATCATCAGCTGGGGCGAGGGCTGTGC 817
267  aGluTrgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSer 284
818  CCAGCGCAACAGCGCCGGGCTTACATCAGCTCTCTGCGCACCGCTCT 867
284  rPValGluLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
868  GGGTGGAGAAAGATCTGCAAGGGGTGACGCTCCGCGGCGGCTCAGGG 917
301  GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
918  GGTGGGGCCCTCAGGGCACCGAGCGAGGCTCTGGGGCGCGCGCTC 967
317  r 317
968  C 968

```

seq_name: gb-pr:BC009726
seq_documentation_block:

LOCUS BC009726 1403 bp mRNA linear PRI 22-OCT-2001
DEFINITION Homo sapiens, protease, serine, 22, clone MGC:9599 IMAGE:3899480,
mRNA, complete cds.
ACCESSION BC009726
VERSION BC009726.1 GI:16307274

KEYWORDS

MGC.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1403)
Strausberg, R.
Direct Submission
Submitted (29-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-sngc.stanford.edu>
Contact: (Dickson, Mark) mcdexaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

FEATURES

source
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAX Plate: 14 Row: 1 Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 11545838.
Location/Qualifiers
1..1403
/organism="Homo sapiens"
/db_xref="locusID:64063"
/db_xref="taxon:9606"
/clone="MGC:9599 IMAGE:3899480"
/cison="type="Pancreas, epithelioid carcinoma"
/clone_lib="NTH_MGC_70"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
39..992
/codon_start=1
/product="protease, serine, 22"
/protein_id="AAH09726.1"
/db_xref="gi:16307275"
/translation="MVSAGAPALGGGGLGFTSLILASTAILNARIIPVPPGCKP
COLNRVVGEDSTDEMPWIVSIQKNGTHHCAGSLITRWVITAHCKRMVLPYLP
SVLIGAMOLGNPGRSOKGVAMVPEPVYEMKESGACADIALVRLERSIQSERVLP
CLPDASTILPPTWHWISGMSIDGVPLOKPLPDIIDSEVCSHLVWKGAGG
GPTIDMICAATLBERDPCIGDSGPTMCOVDAMILAGTIISNGEGCAENRPVYI
SISAHRSVERKIVGVQLRGRAGGALRAFSQSGAARAS"

CDS

BASE COUNT 275 a 452 c 410 g 266 t
ORIGIN
alignment_scores:
Quality: 1720.00 Length: 317
Ratio: 5.426 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-040-803-7 x BC009726 ..
Align seg 1/1 to: BC009726 from: 1 to: 1403

```

1  MetValAlaSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyTh 17
|||||
39  ATGCTGTTTGTGAGAGCCGCCAGCCCTGGGTGGGGGCTGTCTGGCAC 88
|||||
17  rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34
|||||
89  CTCACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 138

```

```

34  rgileProValProProAlaCysGlyLysProGlnInLeuAsnArgVal 50
139  GGATACCTGTTCCCGACCTGGGAGAGCCCGACAGCTGAACCGGGTT 188
51  ValGlyGlyLysAspSerThrAspSerGluTrpProTrrPleValSer1 67
189  GTGGGGGGGAGGAGACAGCTGACAGCGAGTGGCCCTGGATGCTGAGCAT 238
67  eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArgT 84
239  CCAGAGAAATGGGACCCACACAGCTGCGAGTCTGCTGCACACAGCGCT 288
84  rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
289  GGCTGATCTACCTGCTGCCACTGTTTCAAGACACCTGAACCAACCATAC 338
101  LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAspProGlySerAr 117
339  CTCTTCTCTGCTGCTGGGGGCTGGCAGCTGGGGAACCTGGCTCTCG 388
17  gSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSerT 134
89  GTCCCAAGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 438
134  rPlySeluGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
439  GGAAGAGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 488
151  IleGlnPheSerGluArgValLeuProIleGlySerProAspAlaSer1 167
489  ATACATTTCTCAGAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
167  eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySer1IleG 184
539  CCACCTCCCTCCCAACACACACCTGATGATGATGATGATGATGATGATG 588
184  IAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
589  AAGATGAGATGCTCCCTGGCCACCTCAGACCTCAGACCTCAGACCTCAG 638
201  ProIleIleAspSerGluValCysSerHisLeuTyrTrpArgGlyValG1 217
639  CCGATATGAGCTCGGAGAGTCTGCAACCATCTGCTGCTGCTGCTGCTGCT 688
217  yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGluGlyG 234
689  ACAGGAGACCATCAGAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738
34  IuArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnVal 250
139  AGCGGATGCTGCTGCTGGCGAGCTCGGGGGGCTCCCTCATGTGCGAGGTG 788
251  AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAl 267
789  GACGGGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 838
267  aGluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSerT 284
839  CGAGGCAACAGGCGCGGGCTCATATCAGCTCTGCGGACCGCTCTCT 888
284  rPValGluLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
889  GGGTGAAGAAAGATCGTGCAGAGGGGTGACGCTCCGGGGCGCGCTAGGGG 938
301  GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
939  GGTGGGGGCTCAGGAGCAGCAGGAGGCTGTGGGGCGCGCGCGCTC 988
317  r 317
989  C 989
seq_name: gb_AB010778

```

```

seq_documentation_block:
LOCUS      AB010778                1321 bp    mRNA    linear    ROD 17-MAY-2001
DEFINITION Mus musculus mBSP-4 mRNA for brain specific serine protease-4,
complete cds.
ACCESSION  AB010778
VERSION    AB010778.1  GI:12248748
KEYWORDS   Mus musculus
SOURCE     Mus musculus
ORGANISM   Mus musculus brain cDNA to mRNA.
REFERENCE  1 (sites)
AUTHORS    Mitsui,S., Okui,A., Kominami,K. and Yamaguchi,N.
TITLE      Cloning and characterization of a novel serine protease, mBSP-4
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1321)
AUTHORS    Yamaguchi,N. and Mitsui,S.
TITLE      Direct Submission
JOURNAL    Submitted (27-JAN-1998) Nozomi Yamaguchi, Institute for Geriatrics,
Kyoto Prefectural University of Medicine, Department of Cell
Biology, Kajicho 465, Kamigyo-ku, Kawaramachi Hirokoji, Kyoto,
Kyoto 602-8566, Japan (E-mail: nozomi@koto.kyu-u.ac.jp,
tel:81-75-251-5848(ex.5848), fax:81-75-251-5848)
FEATURES
     source
         1..1321
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /tissue_type="brain"
     gene
         80..1000
             /gene="mBSP-4"
             /note="mBSP-4"
             /note="prosemin"
             /codon_start=1
             /product="brain specific serine protease-4"
             /protein_id="BAB20262.1"
             /db_xref="GI:12248749"
             /translation="MMISRPPLAGDGFILILVLTSTAPISAATIRVSPDGRP
OQINRIVGSEDSADQMPVIVSLKNGSHHCAGSLIRNRYVTAHCFKSNMDSLP
SVLGAMKRGSPGSRQKGLAMVLPNRYSWKESTHADILVRLHESIOFSERTLP
CLPDSVRLRPKTDGMIAGWSIDQGVLPHPQRLKVLPIIDSELRKSLYMGAGQ
EATKSGMCAQGLIEGERDACLGSQGLMCQVDHMLTLGIISNEGCAADRPRVYTS
LTAHRSWQRIQGVQLKGYLADSGDTGS"
     BASE COUNT
         280 a      375 c      372 g      294 t
     ORIGIN
         1  MetValValSerGlyAlaProProAlaLeuGlyGlyLysLeuGlyThr 17
           |||:|||||  |||:|||||  |||:|||||  |||:|||||
         80  ATGATGATTCACAGACTCCCGCAGCAGCTGGGTGGGAGCAGTTCAGCAT 129
           |||:|||||  |||:|||||  |||:|||||  |||:|||||
         17  rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34
           |||:|||||  |||:|||||  |||:|||||  |||:|||||
         130  CTATATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 179
           |||:|||||  |||:|||||  |||:|||||  |||:|||||
         34  rgileProValProProAlaCysGlyLysProGlnInLeuAsnArgVal 50
           |||:|||||  |||:|||||  |||:|||||  |||:|||||
         180  CCATCCAGATGTCCTCCAGACTGTGGGAGGCTTCAGCAGCTGAACCGGATT 229
           |||:|||||  |||:|||||  |||:|||||  |||:|||||
         51  ValGlyGlyLysAspSerThrAspSerGluTrpProTrrPleValSer1 67
           |||:|||||  |||:|||||  |||:|||||  |||:|||||
         230  GTGGAGGTGAGGAGCAGCATGATGCCACAGTGGCCCTGGATGTAGCAT 279

```

alignment_scores:

Quality: 1275.50 Length: 312
 Ratio: 4.621 Gaps: 2
 Percent Similarity: 88.462 Percent Identity: 75.641

alignment_block:

US-10-040-803-7 x AB010778 ..

Align seg 1/1 to: AB010778 from: 1 to: 1321

```

1  MetValValSerGlyAlaProProAlaLeuGlyGlyLysLeuGlyThr 17
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
80  ATGATGATTCACAGACTCCCGCAGCAGCTGGGTGGGAGCAGTTCAGCAT 129
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
17  rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
130  CTATATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 179
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
34  rgileProValProProAlaCysGlyLysProGlnInLeuAsnArgVal 50
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
180  CCATCCAGATGTCCTCCAGACTGTGGGAGGCTTCAGCAGCTGAACCGGATT 229
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
51  ValGlyGlyLysAspSerThrAspSerGluTrpProTrrPleValSer1 67
   |||:|||||  |||:|||||  |||:|||||  |||:|||||
230  GTGGAGGTGAGGAGCAGCATGATGCCACAGTGGCCCTGGATGTAGCAT 279

```


seq_name: gb_to:RA5642

seq_documentation_block:

LOCUS RRA5642 1259 bp mRNA linear ROD 07-SEP-1998

DEFINITION Rattus rattus mRNA encoding brain serine protease bsp2, partial.

ACCESSION AJ005642

VERSION AJ005642.1 GI:3559979

KEYWORDS bsp2 gene; serine protease.

SOURCE black rat.

ORGANISM Rattus rattus

Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

```
67 ecglnlyasnglythrshisCysAlaGlySerLeuLeuhsrArgt 84
|
280 CCTCAAAATGGCTCCACCATCTGAGGGCTCCCTGCTACCAACCGCT 329
84 rPvAlIleThrAlaAlaHisCysPheIysAspAnLeuAsnLysProTyr 100
|
330 GGGTGGTCACAGCCGCCACTGCTTTAAGAGCAATATGAAACAAACACT 379
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
380 CTGTTCACAGTATTGTTGGGGCTCGGAAGCTGGGGAGCCACAGCCCAAG 429
117 gSerGlnLysValGlyAlaAlaTrpValGluProHisProValYrSert 134
430 GTCCACAGAAATGAGCAATTCCTGGGTGCTCCATCCACCAAGTATTTCT 479
134 rPlySGlIGlyAlaCysAlaAspIleAlaLeuValArgLeuGlnuArgSer 150
180 GGAAAGAGGGAACCATCCATGCATGTCCTGGTGGCTGGAAACATCC 529
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerIl 167
530 ATMCAGTTCCTGAGCGGATCTCGCCCATCTGCATCCCTGACTCTCTGT 579
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerIleG 184
580 CCGTCTCCCTCCCAAGACCGATCGTGGATTCGCGGGTGGGAACATCC 629
184 InAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuLysVal 200
630 AGATGGAGTGGCCCTGCCCCACCTCAGACCCCTCAGAACTGAAGGTG 679
201 ProIleIleAspSerGlyValAlcysSerHisLeuYrTrpArgGlyAla 217
680 CCCATCATCATGACTCCGAACTCTGCAAAAGCTTGACTGCGGGAGCCGG 729
217 yGlnIlyProIleThrGlnuAspMetLeuCysAlaGlyYrLeuGlnuGly 234
730 TCAGGAAGCCATCAGGAGGGCATGCTGTGCTGTGTTACTCGGAAGGG 779
234 luArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnVal 250
780 AGCGGGATGCTGTCTGGGCGACTCGGGGGTCCCTGATGTGCCAGGTG 829
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlnGlyAla 267
830 GATGACCACTGGCTACTGACTGCATTAACACACTGGGGAGAGGGCGCC 879
267 acGluGAsnArgProGlyValYrTrpIleSerLeuSerAlaHisArgSer 284
880 GGAC..GACCGGCGCGGTGTGTACACGAGCCTCTCAGTCAACGCTCT 926
284 rPvAlIGluLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
927 GGGTGCAAGGATCGTTCAAGGGGTCACTGCGCGGTACTTGCGCGGAC 976
301 Gly.GlyAlaLeuArgAlaProSerGlnGlySer 311
977 AGTGGGGACAGGAGCTCTAATCTAAGATCT 1010
```

[illegible]

```

275 GCTCCACACACTGCGCAGGCTCTGCTACCAACCGCTGGGTGTCTCA 324
88 AlaAlaHisCysPheIysAspAsnIleuSlnysProTyrIlePheSerVa 104
325 GCGGACACTGCTTTAGCAGCATATGACACAAACCATCTCCGTACTCAGT 374
104 IleuLeuGlyAlaTyrPglIleuGlyAsnProGlySerArgSerGlnYsv 121
375 ATTGTTGGGGGCTGTGAGAGTAGAGAACCGAGCCCAAGCTCAGAAAG 424
121 aGlyValAlaTyrValGluProHisProValTyrSerTrpLysGluGly 137
435 TAGGCACTTCTCGGTGCTGCTGCCCAACCCAGATTTCTCGGAGAGAGGG 474
138 AlaCysAlaAspIleAlaLeuValArgLeuGlnArgSerIleGlnPheSe 154
475 ACCCATGAGACATTGCCCTGTGTGGCTGCGAAGCCCAATTCATTCCTC 524
154 rGluArgValIleuProIleCysLeuProAspAlaSerIleHisLeuPro 171
525 TGAGCGGATCCTGCCCATCTGCTACCTGACTCCTGTCATCTCCCTC 574
171 roAsnThrHisCysTrpIleSerGlyTrpGlySerIleGlnAspGlyVal 187
575 CCAACACCACTGCTGATTCGCGGAGGAGACATCCAGATGAGAGTG 624
188 ProIleuProHisProGlnThrLeuGlnIleuSlnysValProIleIleAs 204
625 CCCCTGCCCGCCCTCAGACCCCTCAGAGCTGAGAGTCCCATCATCTCA 674
204 pSerGluValIcySerHisLeuTyrTrpArgGlyAlaGlyGlnGlyPro 221
675 CCTGGAAGCTCTGCAAAAGTTGTACTGGCGGAGCTGTCAGAGAACCA 724
221 IeThrGluAspMetLeuCysAlaGlyTyrIleuGlnGlyAluArgAspAla 237
725 TCACGAGAGACATGCTGTGTGTTACCTGGAAGGAGCGGAGCGCT 774
238 CysLeuGlyAspSerGlyGlyProLeuMetCysGlnValAspGlyAlaTr 254
775 TGTCTGGGCGCAGCTGTGGGGGTCCTGATGTGCCAGGTGATGACCCCTG 824
254 pIleuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAlaGluArgAsnA 271
825 GCYACTGACGGGCAATATCAGCTGGGAGAGGGCTGGCGCAGCAGCAAC 874
271 rGProGlyValTyrIleSerLeuSerAlaHisArgSerTrpValGlnYls 287
875 GGCCCGGCGGTGTACACCAAGCTCTCCTAGCTACCGCCCTGGGTGCAAGG 924
288 IleValGlnGlyValGlnLeuArgGlyArgAlaGlnGlyGly 302
925 ATGCTGCAGAGGGGTGCTGCGAGGGCGCTTGGCGGACAGCTGGG 969
seq_name: gb_pr:AC003965
seq_documentation_block:
LOCUS AC003965 34642 bp DNA linear PRI 31-DEC-1997
DEFINITION Homo sapiens chromosome 16, cosmid clone 325D7, complete sequence.
ACCESSION AC003965
VERSION AC003965.1 GI:2734091
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 34642)
Ricke,D.O., Bruce,D., Muntz,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J. and Deaven,L.
Sequencing of Human Chromosome 16p13.3
JOURNAL
unpublished

```

```

REFERENCE
2 (bases 1 to 34642)
Ricke,D.O., Wagner,R.P. and Muntz,M.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL
Unpublished
3 (bases 1 to 34642)
Ricke,D.O., Bruce,D., Muntz,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J. and Deaven,L.
Direct Submission
JOURNAL
Submitted (31-DEC-1997) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
COMMENT
Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source
1..34642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="16p13.3"
/clone="325D7"
/complement(96..381)
/rpt_family="Alu"
/complement(955..1255)
/rpt_family="Alu"
/complement(1803..2246)
/rpt_family="Alu"
2626..2809
/rpt_family="Alu"
3105..3373
/rpt_family="Alu"
4249..4384
/rpt_family="Alu"
/complement(4506..4797)
/rpt_family="Alu"
6206..6510
/rpt_family="Alu"
/complement(6664..6751)
/rpt_family="Alu"
/complement(8846..9171)
/rpt_family="Alu"
9425..9509
/rpt_family="Alu"
9509..9573
/rpt_family="MIR1b"
9672..9766
/rpt_family="MSTRC"
10069..10364
/rpt_family="MSTRC"
10370..10674
/rpt_family="Alu"
/rpt_family="Alu"
/complement(10884..11179)
/rpt_family="Alu"
11211..11522
/rpt_family="Alu"
11493..11799
/rpt_family="Alu"
/rpt_family="Alu"
/complement(11793..12160)
/rpt_family="L1"
13964..14098
/rpt_family="MIR"
/complement(14456..14468)
/rpt_family="Alu"
16240..16446
/feature="45% Swiss-Prot:PI8127"
join(<17398..17540,17771..18048,19600..19757,20293..20878)
/feature="SP0011A"
/evidence-not_experimental
join(<17398..17540,17771..18048,19600..19757,20293..20529)
/feature="SP0011A"
/feature="putative serine protease"
/codon_start=1

```

```

/evidence=not_experimental
/product="SP001LA"
/protein_id="AAB93671.1"
/db_xref="GI:2734092"
/translation="LNRYVGGSDSTDEMPYIVSIQKNGTHHCAGSLIJSRWVITAAH
CEKDNINRKYLSVLGAWQNGPSRSOKGVANVPHRYPSKEGACADIALYRLE
RSIQSERVLPICLPDASIHLPNTHCWISGWSIQDGVLPFPQTLQKLVPIIDSE
VCSHLYRWGAGGPIEDMIGAGYLEGERDACLQDSGLQKQVDMGAMLAAGIISWGE
GCAERNRPGVYISLSAHSRSMVEKIVGVQLRGRAGGGGALRAPSGSGAAARS"
repeat_region complement(18257..18556)
                /rpt_family="Alu"
polya_site 20873..20878
            /gene="SP001LA"
repeat_region complement(22028..22333)
            /rpt_family="Alu"
repeat_region complement(22376..22722)
            /rpt_family="Alu"
repeat_region complement(24599..25217)
            /rpt_family="Alu"
repeat_region complement(25525..25603)
            /rpt_family="Alu"
repeat_region complement(25703..26005)
            /rpt_family="Alu"
repeat_region complement(26818..26888)
            /rpt_family="LTR8"
repeat_region complement(26927..27263)
            /rpt_family="MER4"
repeat_region 27294..27592
            /rpt_family="Alu"
repeat_region complement(27650..27931)
            /rpt_family="MER4b"
repeat_region complement(29107..29227)
            /rpt_family="Alu"
repeat_region complement(30340..30649)
            /rpt_family="Alu"
misc_feature 32294..32459
            /note="Serine Protease SP002LA catalytic exon A"
misc_feature 32602..32780
            /note="Serine Protease SP002LA catalytic exon B"
misc_feature 33314..33474
            /note="Serine Protease SP002LA catalytic exon C"
misc_feature 33691..33966
            /note="Serine Protease SP002LA catalytic exon D"
BASE COUNT 7787 a 9148 c 8812 g 8895 t
ORIGIN

```

```

Percent Scores:
Quality: 1094.50      Length: 1055
Ratio: 3.895          Gaps: 5
Percent Similarity: 26.635      Percent Identity: 26.635

```

alignment_block:

US-10-040-803-7 x AC003965 ..

Align seg 1/1 to: AC003965 from: 1 to: 34642

```

36 ProValProProAlaCysGlyLysProGlnGlnLeuAsnArgValValG1 52
|||||
17365 CCAGTTCCTCCCAAGCCCTGTGGGAAGCCCAAGCACTGAACGGGCTGTGG 17414
|||||
52 YGUYGUUASPserThrAspSerGluTrpProTrpPileValSerIleGlnL 69
|||||
17415 CGGCGAGGACAGACTGACGACGAGGAGGCGCTGATCGTACGATCCACA 17464
|||||
69 ysaGngLyrThrHisCysAlaGlySerLeuLeuThrSerArgTrpVal 85
|||||
17465 AGAATGGGACCCACACTGCGGAGTTCTGTCTCACGACGCGCTGGGTG 17514
|||||
86 IleThrAlaAlaHisCysPheLys..... 93
|||||
17515 ATCACTGCTCCCACTGTTCAAGAGATATGACAGCCGCGCTGGGCGAC 17564
|||||

```

```

93 ..... 93
17565 TTGATCTTCTAAGGCCCTGAGGCTTGTGTGCCACGAGCCCTGGGCTCCTT 17614
93 ..... 93
17615 ACAGGCTCTGCTTCGGGGCCTGTTCATCTTGTGCCCCCTGATTTCA 17664
93 ..... 93
17665 CCTAGAAGCCTCTCTGCTTTCAGGGGCTTTCACACCATGCTGTTCCC 17714
93 ..... 93
17715 CAGGTACCTCTCCCAAGACGAGCCCTGGCTGATCTCCCTTCCCTTCC 17764
94 ..Asp..AsnLeuAsnLysProTyrLeuPheSerValLeuLeuGlyAlaT 109
|||||
17765 TTCACGACACCTGAAACAAACCATACCTGTTCTCTGCTGCTGGGGGCT 17814
109 rpgInLeuGlyAsnProGlySerArgSerGlnLysValGlyValAlaTTP 125
|||||
17815 GGCAGCTGGGGAAACCTGGCTCTGGTCCAGAGAGTGGGTTCCTGG 17864
126 ValGluProHisProValTyrSerTrpLysGluGlyAlaCysAlaAspI 142
|||||
17865 GTGAGGCCCAACCTGCTGATTCCTGGAAGAGTGCCTGTGCAGACAT 17914
142 eAlaLeuValArgLeuGluArgSerIleGlnPheSerGluArgValLeuP 159
|||||
17915 TGCCCTGCTGCTCTCGAGGCTCCATACAGTCTCAGAGGGGCTCTGC 17964
159 rolleCysLeuProAspAlaSerIleHisLeuProAsnThrHisCys 175
|||||
17965 CCATCTGCTTACCTGATGCTCTTCCACCTCTCCCAACACCCACTGC 18014
176 TrpIleSerGlyTrpGlySerIleGlnAspLys..... 186
|||||
18015 TGATTCACAGCTGGGGAGATCAGATGAGAGTGCAGCTTCTTCTTC 18064
186 ..... 186
18065 TCCAAAGTGTGGGGGACAGATGGGGAAGAGATGCGACACTCAGGG 18114
186 ..... 186
18115 ACACAGACCATGACGACAAAGGTGGAGGCCAGAGGACACATGAC 18164
186 ..... 186
18165 AGAGGTGAAGAAAGATGGGCTGCAATCCAGAAAGAACTTGGTGGCA 18214
186 ..... 186
18215 ATTTCAGGCATTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 18264
186 ..... 186
18265 TTTTCTTGTGACAGAGGTGTACTGTGTCACCAAGCTGAGTGCAGT 18314
186 ..... 186
18315 GGTGTGATCTAACTCACTGAGCCTTGTGCTCCAGAGCTCAAGTATCC 18364
186 ..... 186
18365 TCCCACTCAGCCTCCCAAGTACGTGAAGTGTGTGTACACCA 18414
186 ..... 186
18415 CTTGCTAATTTTGTATTTTGTAGAGACAGGCTTTTCTGTGTGGC 18464
186 ..... 186

```

18465 TGGCTGGTCTTGAAGTGTGGCTCAAGCAATCCACAGCTCAGCCTC 18514
186 186
18515 CCAAGATGCGGAGTTACAGGTGACGCCACTGCTGCTCATTTCT 18564
186 186
18565 ACTCAAGATCTGGCTATGTTTGAAGCAACAGCCAGACCCGTGAAG 18614
186 186
18615 CTATTCATTCAATCAATACAGAAATGGCTGAAGGGGTGTGTGGGAC 18664
186 186
18665 TAAGCTAATACAGGCTGAATATTAAGAGCGCTGATTCAGAGTGGGAA 18714
186 186
18765 GAAATGACGCCACTTAATTGTAAATATTAATTTCAAAACCACTGG 18814
186 186
18815 AGGTTTGGAAAATCATCTCAAAACTCAAACTAGATTACATTAAGCAAG 18864
186 186
18865 AACGAGGAAAGCAGCAAGAAAGTAGATGACAGATCGACAGGCAGATG 18914
186 186
1915 GATGATAGAGCTGGGCCAAGAAACATACAGGAGAAATGAGGA 18964
186 186
1965 TACAATGAAGACCTGGAAGAAAGATTGAAGAGAGAGCTAGAGACTTA 19014
186 186
19015 GTAATAACAAGTGTAGACAGATGAATAATTACAGATTCAGAAATA 19064
186 186
65 TTAAGAGAAATGACACTAGCTAGGTGTGTGGGATCCAGTAACCTGA 19114
186 186
19115 GGTTCACCTCTCCATGATCTGTGAGGGGCTTCTCAAGAGTAGAAT 19164
186 186
19165 CGCGGACTTTCAGTCCACATCTGACACAGCTGTGAGACCACTCC 19214
186 186
19215 AGGCAAGGGTGGGATCCCTCAGGTCCCTGTAATTTCAACATGAGTG 19264
186 186
19265 GACATTTATTGAACACTTCTTATAGCAGTCCAGTGTGATGAAGTA 19314
186 186
19315 ATGCAATCTGAGACAGGAAGGCTTAACCATCTGTGATATTAAGAA 19364
186 186

19365 ATTATTAAAGAGTAGGCTTTAGGTTTCATTTATGTTCTATATCAG 19414
186 186
19415 CAGGGCAGAGGTGAGACCTGAATTTCAATTTCTGAAGCATCTGG 19464
186 186
19465 GAGAGACTGTAGACTAATCTGTTAGCCTTTGAGAACCCCTGCTGAG 19514
186 186
19515 GAATGATTTGGGGGAATAGTCTTTGAGAGCAACGAGAGGGAG 19564
187 187
19565 ACCCCAAACCTGATTCGTCTCTCTCCCTCCAGCTTCCTGGCCCCAC 19614
192 roGlnThrLeuGlnIleuValProIleIleAspSerGluValCys 208
19615 CTCACACCCCTGAGAGAGCTGAAGGTTCTATCATGACTCGGAAGTCTGC 19664
209 SerHisLeuTyrTrpArgGlyAlaGlyGlnGlyProIleThrGluAspMe 225
19665 AGCCATCTGTACTGGCGGGGAGACAGGACAGGACCATCATCTGAGGACAT 19714
225 IleuGlyAlaGlyTyrLeuGlnGlyGluArgAspAlaCysLeu..... 239
19715 GCTGTGCCGCTACTGTGAGGGGAGCGGATGCTGTCTGTGAGCT 19764
239 239
19765 CCTCGAGCCCCCACCCTTGCCAGAGAGGGCTTCGGAAAGAGACGAG 19814
239 239
19815 TCAGTCTGTCCACACTGAGCCCAACAACCTCTCTGCTCTCCCA 19864
239 239
19865 GAATGGCTGTGCTGACAGTCCGGGCTCCCATAGAACAGCCTCAGCC 19914
239 239
19915 TGGCTCAGCACCTCATTATTTGTCAAGACATTAACCTGGGCAATCCAGC 19964
239 239
19965 TGACACACCTGCCCTCCGAGGCTTCAATTCCTCAGGGGCTTCCA 20014
239 239
20015 CGCCTCAGTCCCGCAGAGCAGCTGACTCAGGGGCTGTCTGAAAGCG 20064
239 239
20065 TCAGATGTTGAGGGGAGACACCGCGCTGAGAGTCTTAAGGCTGGCT 20114
239 239
20115 GCACTCTACTCTGCGCGGGGACTTCTACCTGGTGGAGCTCAGGT 20164
239 239
20165 CCTGTCTCTCAGAGAGAGAAAGGGCTCCGGGAGAGGGCTCATCTG 20214
239 239
20215 CCCCCTGACAGGGGAGGCGCTGATTTGCAACCCCTGTGCAAGGG 20264
240 240
20265 GGTTCCTAACCTGCTCTCTCTGAGGGGAGCTCCGGGGGCCCTCTCA 20314

247 etcysglvalaspclalatrpleuleuallaglyllelesertrpily 263
 20315 TGTGGAGGTGGAGCGCGCTGCTGCTGCGCGCATCATGAGCGGGGC 20364
 264 GIUGLYCYVALAGLUAARSAARPROGLYVALTYRILESERLEUSERAL 280
 20365 GAGGGCTGTGCGAGCGCAACAGGCCCGGGGCTCATCATGAGCTCTTGGC 20414
 280 aHisarSerTrpValGluValIleValGlnGlyValGlnLeuArgGlyA 297
 20415 GCACGCGCTCTGCTGGTGGAGAAATCGTGCAGAGGGTGCAGCTCCGCGGAC 20464
 297 rGAlaGlnGlyGlyValAlaLeuArgAlaProSerGlnGlySerGlyAla 313
 20465 GCGCTCAGGGGGGTGGGCGCTCAGAGGCAACGAGCGAGGCTGTGGGCGC 20514
 314 AlaAlaArgSer 317
 315 GCCGCGCGCTCC 20526
 name: gb_pr: HSA306593
 seq documentation block:
 LOCUS HSA306593 1013 bp mRNA linear PRI 31-MAR-2001
 DEFINITION Homo sapiens mRNA for marapsin (MPN gene).
 ACCESSION AJ306593
 VERSION AJ306593.1 GI:13516325
 KEYWORDS marapsin; MPN gene.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1013)
 AUTHORS Fortunato, M., Dando, P. M., Rawlings, N. D. and Barrett, A. J.
 TITLE Cloning, sequencing and expression of marapsin, a human serine
 proteinase
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1013)
 AUTHORS Fortunato, M.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2001) Fortunato M., MRC Molecular Enzymology
 Laboratory, Babraham Institute, Babraham Hall, Babraham, Cambridge,
 CB2 4AT, UNITED KINGDOM
 COMMENT Related entry AM170323
 FEATURES
 source
 1..1013
 /location="Qualifiers"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE 2698030"
 20..892
 /gene="MPN"
 20..892
 /gene="MPN"
 /function="putative trypsin-like activity"
 /codon_start=1
 /product="marapsin"
 /protein_id="CAC35467.1"
 /db_xref="GI:13516326"
 /translation="MKRPAAVPLALLLICFGSORAKATACRPPMLNRNMGGOTQES
 EMPQVYSIORNGSHFCGSLIAEOMVLTAAHCFRNTSETSLYVLLGARQLVQPHH
 MYARVQESNPILYOGTASADVALVELAEVPTNTYILPYCLDPSTVETFGNCFW
 TQMGSEEDLLEPRILQKLAVIDIPKCNLYSKDTEGYPKTKTKMDLCAEGE
 EKKKACKGSGGGLVCLVGSQWLDAGVISWGEQCARQNRPGVYIRVTAHHNHTRRI
 PKLRQPARLGGK"
 BASE COUNT 192 a 334 c 321 g 166 t
 ORIGIN
 alignment_scores: 739.00 Length: 285
 Ratio: 3.421 Gaps: 4
 Percent Similarity: 75.789 Percent Identity: 48.772

alignment block:
 US-10-040-803-7 x HSA306593 ..
 Align seg 1/1 to: HSA306593 from: 1 to: 1013
 21 LeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAlaArgIleProva 37
 47 CTGCTGCTGCTGTTTTGGCTCAGAGCGCCAGCGC 84
 37 LProProAlaCysGlyTrpGlnGlnLeuAsnArgValValGlyGly 54
 85 AGCAACAGCCTGTGTGCGCCAGAGATCGAACCGATGATGGCGGCG 134
 54 LuAspSerThrAspSerGluTrpProTrpIleValSerIleGlnLys 70
 135 AGGACACCGCAGAGGCGAGTGGCGCTGGCAAGTCACATCCAGCGC 184
 71 GlyThrHisCysAlaGlySerLeuLeuThrSerArgTrpValIleTh 87
 185 GGAAGCCACTCTGCGGGGCGAGCTCATCGGAGCAGTGGCTCTGAC 234
 87 rAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyrLeuPheSer 104
 235 GCGTGGCGCAGTCTCCG...AACCTCTGAGACGTCCTGTGACAGG 281
 104 allLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerArgSerGlnLys 120
 282 TCCTGCTGGGGCAGAGGAGCTAGTGCAGCGGAGCACACAGCTATGAT 331
 121 ValGlyValAlaTrpValGluProHisProValTyrSerTrpLysGlu 137
 332 GCCCGGAGGAGGAGTGTGAGAGCAACCCCTGTACAG...GGCAGCGC 378
 137 yAlaCysAlaAspIleAlaLeuValArgLeuGluArgSerIleGlnPhe 154
 379 CTCAGCGCTGACGTGGCTGTGGAGCTGTGAGGAGGACCACTGCTTCA 428
 154 erGluArgValLeuProIleCysLeuProAspAlaSerIleHisLeuPro 170
 429 CCAATTACATCTCCCGGTGGCGCTGACCCCTGATGATTTAG 478
 171 ProAsnThrHisCysTrpIleSerGlyTrpPheSerIleGlnAspGly 187
 479 ACGGCGATGACGTCTGCTGCTGCTGCGGCGACCCCACTGAGAGA 528
 187 lProLeuProHisProGlnThrLeuGlnLysLeuLysValProIleIle 204
 529 CCTCTGCGCGAACCAGGATCTCACAACACTGCTGTGCCATCATCG 578
 204 spSerGluValCysSerHisLeuTyrTrpArgGlyAlaGlyGlnGly 219
 579 ACACACCCAGTGAACCTGCTACAGCAAGACAGCAGATTGGCTTAC 628
 220ProIleThrGluAspMetLeuCysAlaGlyTyrLeuGln 233
 629 CAACCCAAACCATCAGATGACATGCTGTGCGCGGCTTCGAGAGAGG 678
 233 yGluArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGln 250
 679 CAAGAAGATGCTGCTGCAAGGCGAGCTGGGCGGCCCGGTGCTCG 728
 250 aAspGlyValTrpLeuAlaGlyIleIleSerTrpGlyGlnGlyCys 266
 729 TGGGTACGTGCTGTGCGAGCGGGGTGATACCTGGGGGTAGGGCTGT 778
 267 AlaGluArgAsnArgProGlyValTyrTrpIleSerLeuSerAlaHisArg 283
 779 GCCCGCAGAACCGCCAGGTGTCTACATCGCTGTACCGCCACACACA 828
 283 rTrpValGluLysIleValGlnGlyValGlnLeuArgGlyArgAlaGln 300
 829 CTGATCATCGATCATCCCAAACTGACATTCCAGCGAGGAGTTGG 878

300 1ygly 301
|||||
879 GCGGC 883

seq_name: gb_ov:AB018694

seq_documentation_block:

LOCUS AB018694 2078 bp mRNA linear VRT 05-OCT-1999
DEFINITION Xenopus laevis xepsin mRNA for epidermis specific serine protease,
complete cds.

ACCESSION AB018694
AB018694.1 GI:6009514

VERSION epidermis specific serine protease; xepsin.
KEYWORDS xenopus laevis neurula cDNA to mRNA.
SOURCE Xenopus laevis

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodinae; Xenopus.

REFERENCE
1 (sites)
Yamada, K.

The expression control of xepsin by non-axial and planar
posteriorizing signals in Xenopus epidermis
Unpublished (1998)

2 (bases 1 to 2078)
Yamada, K., Takeshima, K. and Takabatake, T.

Direct Submission
Submitted (15-OCT-1998) Kazuto Yamada, Graduate School of Human
Informatics, Nagoya University; Furo-cho, Chikusa-ku, Nagoya
464-8601, Japan (E-mail: yamada@leinfo.human.nagoya-u.ac.jp,
Tel:+81-52-789-2572, Fax:+81-52-789-2567)

Location/Qualifiers
1..2078
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/dev_stage="neurula"
35..1204
/gene="Xepsin"
35..1204
/note="serine protease"
/codon_start=1
/product="epidermis specific serine protease"
/protein_id="BA8494.1"
/db_xref="GI:6009515"

FEATURES
source

gene

CDS

BASE COUNT 603 a 455 c 437 g 583 t
ORIGIN
/translation="MLQYLSFVLIFFIHQACGVYVSNRIYGVGSKRGEMPMQISIS
YKSDICGSLDLSWMTAHGIDSLDVSYYTYIAYOLASADNSTVSRGYSTIK
HPDQYESSGDLIDLEKRPVTPYLPICPSOVORFAGTMCWTGKNTIOEST
PLISFKTIQAKFAVLIIBSVCCGTMESLGIPIPSFIQEDMVCAGKEGIDACQGD
SGGLVCVNNVNWMLQILIVSWGCAEENRPGYTKVQYQDMVCKTWPPLIVSEESP
SVAPSIGSLAPSGPSLIGRGVASTISQTEAGSVNSIETDKNTSTIETEMSM
NNTTMEFFSVSSTISITAIINETKITIDNEAQIHACSLHTIATLIYLFIRFV"
603 a 455 c 437 g 583 t

alignment_scores:

Quality: 661.50 Length: 282
Ratio: 3.227 Gaps: 7
Percent Similarity: 72.695 Percent Identity: 47.163

alignment_block:

US-10-040-803-7 x AB018694 ..

Align seg 1/1 to: AB018694 from: 1 to: 2078

38 PROPR.AlaCysGlyLysProGlnGlnLeuAsnArgValAlaGlyGlyG 54
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
73 CCAACCAAGCCTGGGTACCAAGTAATATCAATATGCGATGTTGGGGCA 122
54 laaspeThrasperGluTrpProTrrPleValSerTleGlnLysasn 70
||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 TGGACAGTAAGAGGGGGAATGGCCTTGGCAATAGAGCTGAGCTACAA 172

seq_name: gb_pat:AX098193

seq_documentation_block:

LOCUS AX098193 1796 bp DNA linear PAT 30-MAR-2001
DEFINITION Sequence 105 from Patent WO0118542.
ACCESSION AX098193
VERSION AX098193.1 GI:13515276

KEYWORDS

SOURCE

ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

71 G|YThrHisHiscysAlaGlySerLeuLeuThrSerArgTrpValIleTh 87
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
173 AGTAAATCTATCTCGCGGGAGATCGCTTACAGACAGCTGGGTGATGAC 222
87 rAlaIleHiscysPheLysAspAsnLeuAsnLysProTrrPlePheSerV 104
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
223 TGGTCTCATCGCATT...GATTCGTTGGATGTTTCATAC...TATACG 266
104 albeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerArgSerGlnLys 120
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 TTTATCTCGGTGATATACAGCTCTGTCGCCGACAAATCTACAGATATCC 316
121 ValGlyValAlaTrpValGluProHisProValTrrSerTrpLysGluI 137
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 AGAGAGGTGAAAAAGTAAACAAAGCACCCAGATTCATAC...GAAGG 363
137 yAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSerTleGlnPhe 154
|::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
364 ATCCAGTGGTGACATTGCTCTATATGAACTGGAGAAACCTGTCACTTTTA 413
154 erGluArgValLeuProLleCysLeuProAspAlaSerTleHisLeuPro 170
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
414 CCGCTTACATCCTACCCATCTGCTCCCTCCCTCAGAGATTCATTTGCA 463
171 ProAsnThrHisCysTrpLysSerGlyTrrPglSerTleGlnAspGlyVa 187
::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
464 GGAGGAGTATGTGCTGGTGACTGATGGGAAATATTCAGAAAGAGAAC 513
187 lProLeuProHisProGlnThrLeuGlnLysLeuValProIleIleA 204
|::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
514 CCCTTTAATCAGCCCTAAACATATTCAGAGAGCACAATTCGTCATCAG 563
204 spSerGluValCysSerHisLeuTrrTPArgGlyAlaGlyGlnGlyPro 220
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
564 ACAGTTCAGTCTGTGGACATGATCCAGATCCAGCTGGCTACATCCCT 613
221lIeThrGluAspMetLeuCysAlaGlyTrrLeuGlnG 233
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
614 GATTCACGCTTCATTCAGAGAGACATGCTGTGTGCTGTTCAAGAGAGG 663
233 yGluArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnV 250
|::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
664 ACGTATGATGCTGCTCCAGGGGAGACTCAGGTGAGACCTCTTGATGATATG 713
250 alAspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlnGlyCys 266
|::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
714 TCACAAATGTGCTGCTTCAGCTCGGTATGTGAGACTGGGATATGCTGTC 763
267 lAlaGluArgAsnArgProGlyValTrrLleSerLeuSerAlaHisArgSe 283
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
764 GCAGAACCAATACATGCTGCTGTATACCAAGTTCAGTATTCACAGGA 813
283 TTrpValGluLysIleValGlnGlyValGlnLeuArgGlyArgAlaGln 300
|::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
814 TTGTTGAAA.....ACCAATGTCCTTTATGTTGTTGAGTAGAG 854
300 1yglylyAlaLeuArgAlaProSerGlnGlySerGlyAlaAla 314
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
855 AAGGACCAAGTGT...GCACCAAGCATTTGACCAAGCATTTGCA 895

REFERENCE 1 (bases 1 to 1796)

AUTHORS Lee, J., Thompson, P. and Lillie, J.
 TITLE Identification, assessment, prevention, and therapy of ovarian cancer
 JOURNAL Patent: WO 0118542-A 105 15-MAR-2001;
 Millennium Predictive Medicine, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..1796
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 327 a 599 c 511 g 359 t
 ORIGIN

alignment_scores:
 Quality: 607.50 Length: 292
 Ratio: 2.993 Gaps: 7
 Percent Similarity: 69.521 Percent Identity: 44.178

seq_block:
 -040-803-7 x AX098193 ..

Align seg 1/1 to: AX098193 from: 1 to: 1796

```

10 LeuGlyGlyGlyCysLeuGlyThrPheThrSerLeu.....Le 23
||||| ||| ||||| : : : |||||
170 CTGGGGCTGGGCACTGGGGCTGGGCTGCTGCTATCTTGGATT 219
23 uleuLaserThralaLeuasnAlaAarglIeproValProProa 40
||||| ||| : : : ||||| : : : |||
220 ACTCGGCTGGGGACAGAGCGGAGGCGGAGGCTCC..... 259
40 laCysGlyLysProGlnLeuasnAargValAlaGlyGlyLysPser 56
||||| ||| : : : ||||| : : : |||
260 .TGGGGTGGTGGCCCCCAAGCA...GCCATCAGAGGTGAGCAGTGA 304
57 ThrAspSerGluTrpProTrpIleValSerIleGlnLysAnglyThrH 73
: : : : : ||||| ||||| : : : : :
305 GTCCGGCTGAGTGGCTGCTGAGTGCATCAGTATGAGGCGTCCA 354
73 HisCysAlaGlySerLeuLeuThrSerArgTrpValIleThrAlaIah 90
| ||||| ||||| : : : ||||| : : : |||
335 TGTGTGTGTGGCTCTCTGCTGCTGAGCAGTGGTGTCTGACGTGCTC 404
90 lscYsPheLysAspAsnLeuasnLysProTyrLeuPheSerValLeu 106
||||| : : : : : ||||| : : : |||
405 ACTGTTCCCGCAGCGAGCACCAAGAA...GCCATGAGGTCAAGCTG 451
107 GAlaIatrpGlnLeuGlyAsnProGlySerArgSerGlnLysValGly 123
||||| ||||| : : : ||||| : : : |||
432 GGGGGCCACAGCTAGCTCTACTCTGAGGAGGAGGAGGTACAGACCT 501
123 lAlatrpAlaGluProHisProValTyrSerTrpLysGlnGlyLacys 140
: : : ||||| ||| : : : ||||| : : : |||
502 GAGGAGCATCATCCCCACCCAGCTAC...CTCCAGAGAGGCTCCCGG 548
140 laAspIleAlaLeuValAargLeuGluAargSerIleGlnPheSerGlu 156
: : : ||||| : : : ||||| : : : |||
549 GGGAGCATGAGCTCTCCACACTGACAGACCACTGCTCCCGGTAC 598
157 ValLeuProIleCysLeuProAspAlaSerIleHisLeuProProAsn 173
: : : ||||| ||||| : : : ||||| : : : |||
599 ATCCCGCCATCTGCTCTCCAGCAGCAAGCCTCTCCCAAGGCGCT 648
173 HisCysTrpIleSerGlyTrpGlySerIleGlnAspGlyValProLeu 190
||||| : : : ||||| : : : ||||| : : : |||
649 CCATGCGACTGCTGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCTG 698
190 rHisProGlnThrLeuGlnLysLeuValProIleLysAspSerGlu 206
||||| : : : ||||| : : : ||||| : : : |||
699 TGACCCCAAGGCTGAGCAGCACTGAGGTGCTGATGATGATGATGAG 748
207 ValCysSerHisLeuTrpArgGlyAla...GlyGlnGlyPro..... 220

```

seq_name: gb_pr:BC001462

seq_documentation_block:

LOCUS BC001462 1809 bp mRNA linear PRI 12-JUL-2001

DEFINITION Homo sapiens, prostate, serine, 8 (prolactin), clone MGC:2133

IMAGE:3138532, mRNA, complete cds.

ACCESSION BC001462

VERSION BC001462.1 GI:12655206

KEYWORDS MGC.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 1809)

Strausberg, R.

Direct Submission

Submitted (12-DEC-2000) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgaabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Sequencing Center (MISC),

Galthersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nrl.nih.gov

Sherchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantirpop, S., Thomas, P.J.,

Tiongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IMAGE Plate: 4 Row: 3 Column: 3

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 1143193.

Location/Qualifiers

FEATURES

source 1..1809
 /organism="Homo sapiens"
 /db_xref="locusID:5652"
 /db_xref="taxon:9606"
 /clone="MGC:2133 IMAGE:3138532"

05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234224.
 PR 25-SEP-2000; 2000US-0234297.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246509.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 08-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259676.
 PR
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-476222/51.
 XX P-PSDB; AAU17037.
 XX
 PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with aberrant expression or
 PT activity of polypeptides, for treating blood clotting disorder,
 PT haemophilia
 XX
 PS
 XX Claim 1; SEQ ID No 134; 601bp; English.
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunoassay assays (ELISAs). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present

```

|||||
27 ATGGTGGTTCTGAGGCGCCCGAGCCCTGGGTGGGGGCTCTCGGCAC 76
17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34
77 CTTGACCTCCCTGCTGCTGGGTGGAGCAGCATCTCAATGCGGCCA 126
34 rGlieProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
127 GGATACCTGTTCCCGACCTGTGGAGGAGCCGACAGCATGAAACCGGTT 176
51 ValGlyGlyLeuAspSerThrAspSerGluTrpProTrpIleValSer11 67
177 GTGGGCGCGAGAGACAGCAGCAGTACAGCAGTGGCCCTGATCGTGACAT 226
67 eGlnLysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerArgT 84
227 CCAGAAAGATGGGACCCACACCTGCGAGGTCTCTCTCCACAGCCGCT 276
84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
277 GGGTATCATCTGCTGCCACCTGTTCAAGGACATGAAACCAACCATAC 326
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
327 CTGTTCTCTGCTGCTGGGGGCTGGCAGCTGGGGAACCTGGCTCTCG 376
117 gSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSer 134
377 GTCCAGAAAGGTGGGTGTTGCTGGGTGGAGCCGACCCCTGTGATTCCT 426
134 rPlyGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
427 GGAAGAGAGGTGCTGTCAGACATTCCTGCTGCTGACGCTCC 476
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSer11 167
477 ATACAGTTCTCAGAGCGGTCTGCCATCTGCTACCTGATGCTCTAT 526
167 eHisLeuProAsnTrpHisCysTrpIleSerGlyTyrPlySer11eG 184
527 CCACCTCCCTCCAAACACCACTGCTGATCTCAGGCTGGGGAGCATCC 576
184 lAspGlyValProLeuProHisProGlnTrpLeuGlnLysLeuVal 200
577 AAGATGAGTCTCCTGCTGCCACCTCAGACCTCAGAACCTGAAGGTT 626
201 ProIleIleAspSerGluValCysSerHisLeuTyrTrpArgGlyAlaG1 217
627 CCTATCATCTGACTCGAACTCTGCAACCATCTGTACTGCGGGAGCAGG 676
217 yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGlnGly 234
677 ACAGGAGACCATCATCTGAGACATGCTGTGCTGCTACTTGGAGGGG 726
234 lArgAspAlaCysLeuGlyAspSerGlyProLeuMetCysAla1a 250
727 AGCGGATGCTTCTGCGGAGCTCGGGGGCCCTCATGTGCGAGG51G 776
251 AspGlyAlaTrpLeuAlaGlyIleIleSerTrpGlyGlnGlyCysAl 267
777 GAGGCGCCCTGCTGTGCGCGGACATCATCTGAGGAGAGGCTGTC 826
267 aGluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSerT 284
827 CGAGCGCAACAGGCCGCGGTCTACATCAGCTCTCTGCGCACGCTCTCT 876
284 rPValGlnLysIleValGlnGlyValGlnLeuArgGlyValArgAlaGlnGly 300
877 GGGTGGAAAGATCTGCAAGGGGTCTCAGCTCCGGGGGCGCTCAGGGG 926
301 GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
|||||

```

927 GGTGGGCGCTCTCAGAGGACCGAGCAGGCGTCTGGGGCGCGCGGCTC 976

317 r 317

977 C 977

seq_name: /STD1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAx87259

seq_documentation_block:

ID AAX87259 standard; cDNA; 1378 BP.

XX AAX87259;

XX 27-SEP-1999 (first entry)

XX cDNA clone encoding human PRO343, amplified in tumour cells.

XX PRO343; UNQ302; cancer; tumour; diagnosis; therapy; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX sig-peptide

XX mat-peptide

XX W09935170-A2.

XX 15-JUL-1999.

XX 05-JAN-1999; 99MO-US00106.

XX 20-NOV-1998; 98US-0109304.

XX 05-JAN-1998; 98US-0070440.

XX 29-APR-1998; 98US-0083500.

XX 10-JUN-1998; 98US-0086414.

XX 10-NOV-1998; 98US-0107783.

XX (GETH) GENENTECH INC.

XX Botstein D, Goddard A, Gurney AL, Hillan KJ, Lawrence DA;

XX Roy MA, Wood WT;

XX WPI; 1999-430385/36.

XX P-PSDB; AAY06482.

XX Antibody against proteins expressed in neoplastic cells, useful for

XX tumor diagnosis and treatment

XX Example 1; Fig 11; 162pp; English.

XX This is the nucleotide sequence of cDNA clone DNA43318 (ATCC 209481)

XX coding for human PRO343 (UNQ302) (see AAY06482). The clone was

XX isolated from a foetal lung library. Amplification of DNA43318

XX (chromosome 16) was observed in primary lung and primary colon

XX tumours, suggesting an association with tumour formation or growth.

XX Antagonists (e.g. antibodies) directed against PRO343 may have

XX utility in cancer therapy. The invention identifies 14 genes (see

XX CC AAX87254-67) that are amplified in the genome of tumour cells. Such

XX CC amplification is expected to be associated with overexpression of

XX CC the gene product and to contribute to tumorigenesis. The encoded

XX CC proteins (see AAY06477-90) may be useful targets for the diagnosis

XX CC and/or treatment (including prevention) of certain cancers, and may

XX act as predictors of the prognosis of tumour treatment.

XX Sequence 1378 BP; 235 A; 461 C; 412 G; 270 T; 0 other;

XX alignment_scores:

Quality: 1720.00 Length: 317
 Ratio: 5.426 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-040-803-7 x AAX87259 ..

Align seg 1/1 to: AAX87259 from: 1 to: 1378

```

1 MetValValSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyTh 17
  |||||
53 ATGGTGTCTTCTGAGGCGCCCGACCGCTGGGGGCTCTCGGAC 102
  |||||
17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAlaAla 34
  |||||
103 CTTCACCTCCCTGCTGCTGCTGGCGTGCAGACGATCCTCAATGCGGCA 152
  |||||
34 rGlyLeuProValProProAlaCysGlyLysProGlnGlnLeuAla 50
  |||||
153 GGATACCTGTTCCCGACCGCTGTGGGAAGCCCGACGAGCTGAACCGGGTT 202
  |||||
51 ValGlyGlyLysAspSerThrAspSerGluTrpProThrIleValSer 67
  |||||
203 GTGGGGCGGAGAGACACACTGACAGCGAGTGGCTGATGCTGAGCAT 252
  |||||
67 eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArg 84
  |||||
253 CCAGAGAAATGGGACCCACACACTGCCAGTCTCTGCTCACACAGCGCT 302
  |||||
84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAlaLysPro 100
  |||||
303 GGGTGTATCTGCTGCCACTGTTTCAAGGACACCTGAACAAACCATATC 352
  |||||
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySer 117
  |||||
353 CTGTCTCTGTGCTGTGGGGCTGGCGCTGGGGAACCTGGCTCTCG 402
  |||||
117 gSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSer 134
  |||||
403 GTCCCGAAGGTTGGTGTGCTGGGTGAGCCCGACCTGTATTCCT 452
  |||||
134 rPlySGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuLysArg 150
  |||||
453 GGAAGGAAGTGCCTGTGAGACATTCCTGCTGCTGAGCGCTCC 502
  |||||
151 IleGlnPheSerGlyArgValLeuProIleCysLeuProAspAlaSer 167
  |||||
503 ATACAGTTCTCAGAGCGGGTCTGCCCATCTGCTACCTGATGCTCTAT 552
  |||||
67 eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpIleSer 184
  |||||
553 CCACCTCCCTCCAAACCCACCTGCTGATCTCAGCGTGGGGGAGCAATCC 602
  |||||
184 InAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
  |||||
603 AAGATGGAGTTCCCTTGGCCCACTCCAGACCTCGACAAACATGAAGTT 652
  |||||
201 ProIleIleAspSerGlyValCysSerHisLeuYrTrpArgGlyAla 217
  |||||
653 CCTATCATCATGACTCGAAGTCTGCAGCATCTGACTGGCGGAGCAGG 702
  |||||
217 YGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrIleuGlnGly 234
  |||||
703 ACAGGGACCCATCATCTGAGACATGCTGTGCGGCTACTTGAAGGGG 752
  |||||
234 LuArgAspAlaCysLeuGlyAspSerGlyLysProLeuMetCysGlyVal 250
  |||||
753 AGCGGAGATGTTGTCTGGGCACTCCGGGGGCCCCCATATGTCGACAGTG 802
  |||||
251 AspGlyAlaTrpLeuLeuAlaGlyTyrIleSerTrpGlyGlnGlyCysAl 267
  |||||
803 GACGGCGCTGGTGTGGCGGCGCATCATCATGCTGGGGCGAGGCTGTGC 852
  |||||

```

```

seq_name: /SID1/9cgcdata/geneseq/geneseqn-emb1/NA1999.DAT: AAX52262
seq_documentation_block:
ID AAX52262 standard; DNA; 1378 BP.
XX
AC AAX52262;
XX
DT 25-JUN-1999 (first entry)
XX
DE Protein PRO343 cDNA clone DNA43318-1217.
XX
KW Secreted protein; transmembrane protein; human; enterocolitis;
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW congenital microvillus atrophy; skin disease; cell growth;
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophid areata;
KW anti-thrombotic; wound healing; tissue repair; ss.
XX
OS Homo sapiens.
XX
WO9914328-A2.
XX
PD 25-MAR-1999.
XX
PF 16-SEP-1998; 98WO-US19330.
XX
XX
25-NOV-1997; 97US-0066840.
PR 17-SEP-1997; 97US-0059113.
PR 17-SEP-1997; 97US-0059115.
PR 17-SEP-1997; 97US-0059117.
PR 17-SEP-1997; 97US-0059119.
PR 17-SEP-1997; 97US-0059121.
PR 17-SEP-1997; 97US-0059122.
PR 17-SEP-1997; 97US-0059184.
PR 18-SEP-1997; 97US-0059263.
PR 18-SEP-1997; 97US-0059266.
PR 15-OCT-1997; 97US-0062125.
PR 17-OCT-1997; 97US-0062285.
PR 17-OCT-1997; 97US-0062287.
PR 21-OCT-1997; 97US-0063486.
PR 24-OCT-1997; 97US-0062814.
PR 24-OCT-1997; 97US-0062816.
PR 24-OCT-1997; 97US-0063045.
PR 24-OCT-1997; 97US-0063120.
PR 24-OCT-1997; 97US-0063121.
PR 24-OCT-1997; 97US-0063127.
PR 24-OCT-1997; 97US-0063128.
PR 27-OCT-1997; 97US-0063329.
PR 27-OCT-1997; 97US-0063327.
PR 28-OCT-1997; 97US-0063341.
PR 28-OCT-1997; 97US-0063542.
PR 28-OCT-1997; 97US-0063544.
PR 28-OCT-1997; 97US-0063549.
PR 28-OCT-1997; 97US-0063550.
PR 28-OCT-1997; 97US-0063564.
PR 29-OCT-1997; 97US-0063435.

```

PR 29-OCT-1997; 97US-0063704.
 PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063735.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-0064248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0063846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0066120.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066772.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0066511.
 PR 24-NOV-1997; 97US-0066453.

(GERTH) GENENTECH INC.

XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 FI WPI; 1999-229533/19.
 DR P-PSDB; AAY13391.
 DR

XX New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 FT

XX Claim 2; Fig 97; 320p; English.

XX AAX52213-74 encode secreted and transmembrane human proteins, and are
 CC obtained from cDNA libraries, prepared from fetal lung, fetal kidney,
 CC fetal brain, fetal liver and fetal retina. The encoded polypeptides
 CC have specific uses based on their homology to known polypeptides.
 CC e.g. PRO217 can be used for disorders associated with the
 CC preservation and maintenance of gastrointestinal mucosa and the repair
 CC of acute and chronic mucosal lesions (e.g. enterocolitis,
 CC Zollinger-Ellison syndrome, gastrointestinal ulceration and congenital
 CC microvillus atrophy), skin diseases associated with abnormal
 CC keratinocyte differentiation (e.g. psoriasis, epithelial cancers such as
 CC lung squamous cell carcinoma of the vulva and gliomas), potent effects on
 CC cell growth and development, diseases related to growth or survival of
 CC nerve cells including Parkinson's disease, Alzheimer's disease, ALS,
 CC neuropathies or cancer. PRO265 can be used as for fibromodulin, e.g. for
 CC reducing dermal scarring. PRO264 can be used as a target for anti-tumor
 CC drugs. PRO533 may be used in the treatment of Usher Syndrome or Atrophica
 CC areata. PRO269 can be used as an anti-thrombotic agent. PRO287
 CC polypeptides and portions may have therapeutic applications in wound
 CC healing and tissue repair. PRO317 can be used for treating problems of
 CC the kidney, uterus, endometrium, blood vessels, or related tissue, e.g.
 CC in the heart of genital tract.
 XX

SO Sequence 1378 BP; 235 A; 461 C; 412 G; 270 T; 0 other;

alignment_scores:
 Quality: 1720.00 Length: 317
 Ratio: 5.426 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-040-803-7 x AAX52262 ..

Align seg 1/1 to: AAX52262 from: 1 to: 1378

1 MetValValSerGlyAlaProProlaLeuGlyGlyGlyCysLeuGlyTh 17
 |||||
 53 ATGGTGGTTTCTGAGCGCCCGCCCTGGGCTGCTGCTGGCAC 102
 |||||
 17 rphrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34

|||||
 103 CTTACCTCCCTGCTGCTGCTGGGCTGACAGCATCTCTAATGGCGCA 152
 34 rglleProValProProAlaCysGlyLysProGlnGlnLeuAsnAlaVal 50
 |||||
 153 GGATACCTGTTCCCGCCAGCTGTGGAAAGCCCGAGCAAGCTGAACCGGGTT 202
 51 ValGlyGlyCysAspSerThrAspSerGluTrpProTrpIleValSer1 67
 |||||
 203 GTGGCGCGCGAGACACACTGACAGCGAGTGGGCTGATCTGTAGCAT 252
 67 eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArgT 84
 253 CCAAGAAATATGGACCCACCACTGCGAGGTCTCTGACACAGCGCGT 302
 84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
 303 GGTGTATCACTGCTGCCACACTGTTTCAAGGCAACCTGAACAAACATAC 352
 101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
 |||||
 353 CTGTCTCTGTGCTGTGGGCTGCGAGCTGGGGAACCTGGCTCTGCT 402
 117 gSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSerT 134
 |||||
 403 GTCCAGAAAGTGGGTGTTCCTGGGTGAGCCCGCCCTGTATTCCT 452
 134 rPLysGlnGlyAlaCysAlaAspIleAlaLeuValAlaGluGluLysSer 150
 453 GGAAGAAAGTGGCTGTGTCAACATGTGCTGTGGTCTGTGAGAGCTTCC 502
 151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSer1 167
 |||||
 503 ATACAGTTTCTAGAGCGGGTCTGCTGCCATCTGCTACTGATGCTCTAT 552
 167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTyrGlySerIleG 184
 553 CCACCTCCCTCAACACCCACTGCTGATCTCAGCGCTGGGAGATCC 602
 184 lAspGlyValProLeuProHisProGlnTrpLeuGlnLysVal 200
 603 AAGATGGAATTCCTTGGCCCACTCAGACCTCAGACCTCAGACACTAAGGT 652
 201 ProlIleAspSerGluValCysSerHisLeuTyrTrpArgIleAlaG1 217
 |||||
 653 CCTATCATGACTCGGAAGTGTGACCATCTGTACTGGCGGAGCAGG 702
 217 yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGlnGly 234
 |||||
 703 ACAGGACCCATCACTGAGAGCATCTGTGCGCGCTACTTGGAGGGGG 752
 234 lAspAspAlaCysLeuGlyAspSerGlyLysProLeuMetCysGlnVal 250
 753 AGCGGATCTGTCTGTGGCGACTCGGGGGGCCCTCATGTGCAGAGTG 802
 251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlnGlyCysAl 267
 |||||
 803 GACGGCGCTGTGCTGTGGCGCATCATCACTGTGGCGAGGCGTGTGG 852
 267 aglUArgAsnArgProGlyValTyrTrpIleSerLeuSerAlaHisArgSerT 284
 |||||
 853 CGAGCCCAACAGGCGCGGGGTCTACTACGCTTCTCTGCGACCGCTCT 902
 284 rPValGlnLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
 |||||
 903 GGGTGGAGAAAGATCGTCAAGGGGTGCACACTCCGGGGCGCGCTCAGGGG 952
 301 GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
 |||||
 953 GTTGGGGCTCTCAGGGACCGAGCCAGGGCTCTGTGGGGCGCGCGGCTC 1002
 317 r 317
 |

1003 C 1003

seq_name: /SIDSL/gcgdata/geneseq/geneseq_emb1/NA2000.DAT:AAA46914

seq_documentation_block:

ID AAA46914 standard; cDNA; 1378 BP.

XX AAA46914;

XX 03-OCT-2000 (first entry)

DE cDNA encoding novel polypeptide PRO343.

XX PRO201; PRO292; PRO327; PRO1265; PRO344; PRO343; PRO347; PRO357;

KW PRO715; PRO1017; PRO1112; PRO509; PRO853; PRO882; tumour cell;

KW tumorigenesis; cancer; neoplastic cell growth; cell proliferation; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

CDS 53..1007

/*tag= a

WO200037640-A2.

XX 29-JUN-2000.

XX 16-DEC-1999; 99WO-US30095.

XX 22-DEC-1998; 98US-0113296.

XX 08-MAR-1999; 99WO-US05028.

XX 02-JUN-1999; 99WO-US12252.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21090.

XX 30-NOV-1999; 99WO-US28313.

XX 30-NOV-1999; 99WO-US28409.

XX 01-DEC-1999; 99WO-US28301.

XX 02-DEC-1999; 99WO-US28565.

XX (GETH) GENENTECH INC.

XX Botstein D, Goddard A, Gurney AL, Hillan K, Lawrence DA, Roy MA;

XX Wood WT;

XX WPI: 2000-452188/39.

XX P-PSDB; AAY93689.

XX New anti-polypeptide antibody useful in the treatment and diagnosis of

neoplastic cell growth and proliferation -

Claim 50; Fig 11; 220p; English.

XX The present sequence encodes a novel human polypeptide. The

XX specification describes novel polypeptides designated PRO201, PRO292,

XX PRO327, PRO1265, PRO344, PRO343, PRO347, PRO357, PRO1017,

XX PRO1112, PRO509, PRO853 and PRO882. These genes are amplified in

XX the genome of tumour cells. The polypeptides are believed to contribute

XX to tumorigenesis. The polypeptides are useful target for the

XX identification of certain cancers, and may act as predictors of the

XX prognosis of tumour treatment. Antibodies against these polypeptides

XX are useful in the treatment and diagnosis of neoplastic cell growth

XX and proliferation in mammals.

XX Sequence 1378 BP; 235 A; 461 C; 412 G; 270 T; 0 other;

XX

alignment_scores:

Quality: 1720.00

Ratio: 5.426

Percent Similarity: 100.000

Length: 317

Gaps: 0

Percent Identity: 100.000

alignment_block:

US-10-040-803-7 x AAA46914

Align seg 1/1 to: AAA46914 from: 1 to: 1378

```

1 MetValValSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyTh 17
|||||
53 ATGGTGTTTCTTGAGAGCCGCCACGCCCTGGGTGGGGCTGTCTCGCAC 102
17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleuSnaIaIa 34
|||||
103 CTTCACCTCCCTGCTGCTGGCTGCGACACACCTCATCTCATGGGCCA 152
34 rGileProValProProAlaCysGlyLysProGlnGlnLeuSnaIaVal 50
|||||
153 GGATACCTGTTCCCCACGCTTGGGAAGCCCACTGACCTGACCGGGTT 202
51 ValGlyGlyGluAspSerThrAspSerGluTrpProTyrIleValSer 67
|||||
203 GTGGCGGCGAGACAGACAGCAGCAGAGTGGCCCTGATCGTGAAGCAT 252
67 eGlnLysAsnGlyThrHisCysAlaGlySerLeuLeuThrArgT 84
|||||
253 CCAGAGAAATGGGACCCACACCTGGCGAGTCTCTCTCACACGCGCT 302
84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
|||||
303 GGGTGATCATCTGCTGCCACCTGTTTCAAGAGCAACCTGAACCAATAC 352
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
353 CTGTTCTCTGCTGCTGCTGGGCGCTGGCAGCTGGGAACCTGGCTCGG 402
117 gSerGlnLysValGlyValAlaTrpValGlnProHisProValLysSerT 134
403 GTCCAGAAAGTGGGTGTGTGCTGGGTGAGACCCACCTGTGTATTCCT 452
134 rPlyGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
453 GGAAGGAAGTGGCTGTGTCAGACATGTCCTGGTGGCTGTGACGCTGCC 502
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSer 167
503 ATACAGTTCAGAGCGGCTGCTCCCATCTGCTCATCTGATGCTCTAT 552
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerIleG 184
553 CCACCTCCCTCAAAACACCACTGCTGATCTCAGGTGGGAGCATCC 602
184 LnaSPGlyValProLeuProHisProGlnThrLeuGlnLysVal 200
603 AAGATGAGTTCCTGCTGCCACCTCAGACCTGCAAGAGCTGAAGGTT 652
201 ProIleLeuAspSerGluValCysSerHisLeuTrpTrpArgLysAla 217
653 CTVATCATCTGAGCTGGAAGTCTGACAGCATCTGTACGGGGAGAGAGG 702
217 yGlnGlyProIleThrGlnAspMetLeuCysAlaGlyTyrLeuGlnGly 234
703 ACAGAGACCATCATCTAGACATGCTGTGCTGGCTACTTGGAGGGGG 752
234 LnaArgAspAlaCysLeuGlyAspSerGlyLysProLeuMetCysGlnVal 250
753 AGCGGATGCTGTGCTGGCGAGCTCCGGGGCCCTCATGTGCCAGGNG 802
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlyCysAl 267
803 GACGGCGCTGCTGCTGGCGGAGCATCATCAGCTGGGAGGCGCTGC 852
267 aGluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSer 284
853 CGAGCGGAACAGGCCCGGGGTCTACATCAGCTCTGTGGCAGCGCTCT 902
284 rPValGlnLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
|||||

```



```

217 yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGlnGly 234
|||||
703 ACAGGAGCCACTCACTGAGGACATGCTGTGCGGCTACTTGAGGGGG 752
224 luArGAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnVal 250
|||||
753 AGCGGAGATGTTGTCGGGCGACATCCGGGGGCCCCCTCATGTGCCAGGTG 802
251 AspGlyAlaTrpLeuLeuAlaGlyTyrLeuSerTrpGlyGlnGlyCysAl 267
|||||
803 GACGGCGCTGCTGCTGCGCGCATCATGAGCGGCGCGGCTGTGC 852
267 agluArGAspArGProGlyValTyrIleSerLeuSerAlaHisArgSer 284
|||||
853 CGAGGGCAAGAGCGCGGCTCATCATGCTGTGCGGCTGTGC 902
284 rPValGlnValSIIeValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
|||||
903 GGGTGGAGAGATGCTGCAAGGGGTGCAGCTCCGCGGCGCGCTCAGAGGG 952
301 GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgse 317
|||||
953 GGTGGGCGCTCAGAGGACCGAGCGAGCGCTGTGGGGCGCGCGCTC 1002
317 r 317
1003 C 1003

seq_name: /SDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAD02990
seq_documentation_block:
ID AAD02990 standard; cDNA; 1430 BP.
AC AAD02990;
XX
XX 31-MAY-2001 (first entry)
XX
DE Human serine protease, protease C-E cDNA.
XX
XX Human; serine protease; protease C-E; therapy; desquamation; skin care;
XX laundry detergent; shampoo; cleaning agent; hair care; skin flaking;
XX neurodegenerative disorder; dermatological; immunogenic; proteolytic;
XX chromosome 16p13.3; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 113..1066
XX /tag= a
XX /product= "Human serine protease, protease C-E"
XX
XX WO200116288-A2.
XX
XX 08-MAR-2001.
XX
XX 14-AUG-2000; 2000WO-US22117.
XX
XX 31-AUG-1999; 99US-0386629.
XX
XX (ORTH ) ORTHO-MCNEIL PHARM INC.
XX
XX Darrow A, Qi J, Andrade-Gordon P;
XX
XX WPI; 2001-226681/23.
XX
XX P-PSDB; AAY72890.
XX
XX Novel serine protease termed protease C-E, useful for treating and
XX preventing skin flaking or imbalance of desquamation -
XX
XX Claim 2; Fig 1; 78pp; English.
XX
XX The present sequence is a human serine protease, protease C-E cDNA which

```

CC is a member of the S1 serine protease family. Protease C-E gene is
 CC located on chromosome 16p13.3 and is expressed in pancreas, placenta,
 CC prostate, small intestine, stomach, spleen, fibroblasts, epidermis,
 CC cerebellum, cerebral cortex, pituitary and hippocampus. Protease C-E is
 CC useful for treating an imbalance of desquamation, by topical application.
 CC A non-pharmaceutical composition comprising the protein may be formulated
 CC as a laundry detergent, shampoo, hard surface cleaning composition, dish
 CC care cleaning composition, skin care composition and hair care
 CC composition. Protease C-E is useful for treating and preventing skin
 CC flaking, neurodegenerative disorders and dermatological pathologies. It
 CC is less immunogenic to sensitive individuals and it provides efficient
 CC proteolytic activity in a non-natural environment.

Sequence 1430 BP; 240 A; 484 C; 428 G; 278 T; 0 other;

alignment_scores:

Quality: 1720.00 Length: 317
 Ratio: 5.426 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-040-803-7 x AAD02990

Align seg 1/1 to: AAD02990 from: 1 to: 1430

```

1 MetValAlaSerGlyAlaProProAlaLeuGlyGlyGlyCysLeuGlyTh 17
|||||
113 ATGGTGGTTCGTGGAGCGCGCCAGCCCTGGGTGGGGCTGTCTGGGCAC 162
17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuSnpAlaAla 34
|||||
163 CTTCACCTCCCTGCTGCTGCTGCGTGCAGACCATCTCATAGGGGCA 212
34 rGleProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
|||||
213 GGATACCTGTTCCCGCAGCTGTGGAGGCCCGACACTGAACCGGTT 262
51 ValGlyGlyGlyAspSerThrAspSerGluTrpProIleValSerI 67
|||||
263 GGGGCGCGCGAGACAGACAGACAGCGAGCGCGCTGAGTCAGACAT 312
67 eGlnLysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerArgT 84
|||||
313 CCAGAAAGATGGAGACCAACACAGCTGCTGTCTGCACGACCGCT 362
84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
|||||
363 GGGTATCATGCTGCCCACTGTTTCAAGGACAACTGAACACCATATC 412
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlySnpProGlySerAr 117
|||||
413 CTGTTCTGTGCTGCTGCGGCGCTGGGAGCTGGGAAACCTGGCTCG 462
117 gSerGlnLysValGlyAlaTrpValGluProHisProValTyrSerT 134
|||||
463 GTCCAGAAAGGTGGTGTGCTGGGTGGAGCCGACCTGTATTCCT 512
134 rPylsGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
|||||
513 GGAAGGAAGGTGCTGTGCAGACATTCCTGCTGCTGCAGCGCTCC 562
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167
|||||
563 ATACAGTTCTCAGAGCGGCTGCTGCCATCTGCTACTGATGCTCTAT 612
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTyrGlySerIleG 184
|||||
613 CCACCTCCCTCCAAACACCACTGCTGATCTCAGGTGGGAGACATCC 662
184 lnaSpGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
|||||
663 AAGATGAGATTCCCTGCCCCACCTCAGACCTGAGAAAGTGAGGTT 712

```

201 ProIleIleAspSerGIValAlCysSerHisLeuYTrpArgGIValAgl 217
|||||
713 CCTATCATGATCGAGTCTGACCATCTGTACTGCGGGGACAGG 762
217 YGIingIProIleThrGIuAspMetLeuCysAlaGIYTrpLeuGIuGIYg 234
|||||
763 ACAGGACCCATCATCTGACGACATGCTGTGCGGGCTACTTGGAGGGG 812
234 IuArgAspAlaCysLeuGIYAspSerGIYGIYProIleuMetCysGIuVal 250
|||||
813 AGCGGATGCTGTCTGTGGGACTCCGGGGGCCCCCTCATGTGCGAGGTG 862
251 AspGIYAlaTrpLeuAlaGIYIleIleSerTrpGIYGIuGIYCysAl 267
|||||
863 GACGGCCCTGGCTGGCTGGCCGATCATCATGCTGGGGGCGAGGGCTGTGC 912
267 AGIuArgAsnArgProGIYAlaTrpIleSerLeuSerAlaHisArgSer 284
|||||
13 CGACGGCAGACAGGCGCGGGCTTACATCAGCCCTCTGCGCAGCGGCTCT 962
284 rPValGIuYsIleValGIuGIYAlaGIuArgGIYArgAlaGIuGIY 300
|||||
963 GGGTGGAGAGATGCTGCAAGGGGTGCGCTCCGGGGCGCGCTCAGGGG 1012
301 GIYGIYAlaLeuArgAlaProSerGIuGIYSerGIYAlaAlaIuArgSe 317
|||||
1013 GTTGGGGCCCTCAGGCGACCGAGCCAGGCTCTGGGGCGCGCGGCTCTC 1062
317 r 317
1063 C 1063

seq_name: /SIDS1/gcgdata/geneseq/geneseqn_emb1/NA2001A.DAT.AAS26871

seq_documentation_block:

ID AAS26871 standard; cDNA; 1439 BP.

AC AAS26871;

DT 07-NOV-2001 (first entry)

DE Human cDNA encoding a novel secreted protein, SEQ ID 63.

XX Human; immunosuppressive; antiarthritic; ss: antirheumatic;

XX cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;

XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

XX vulnary; secreted protein; rheumatoid arthritis;

XX hyperproliferative disorder; cardiovascular disorder; cardiac arrest;

XX cerebrovascular disorder; cerebral ischemia; angiogenesis;

XX nervous system disorder; Alzheimer's disease; infection; ocular disorder;

XX corneal infection; wound healing; epithelial cell proliferation;

XX skin ageing; food additive; preservative; antiproliferative.

OS Homo sapiens.

XX WO200155441-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01320.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239933.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

|||||
654 AAGATGAGTTCCTTCCCTCCACCTCAGACCTCGACAGAGCTGAGGTT 703
201 ProIleIleAspSerGIuValGysSerHisLeuTYrTriArgIValaGI 217
|||||
704 CCTATCATCTGAGCTGGAGTCTGAGCATCTGTACTGGCGGGAGCAGG 753
217 YGIInGIYProIleThrGIuAspMetLeuCysaIaGIYTYrLeuGIuGIY 234
|||||
754 ACAGGAGCCATCTACTGAGGACATGCTGTGCGCGCTACTTGAGAGGG 803
234 IuArgAspAlaCysLeuGIYAspSerGIYGIYProIleuMetCysGIInVal 250
|||||
804 AGCGGATGCTTGTCTGCGGACTCCGCGGGCCCCCTCATGTCCAGGTG 853
251 AspGIYATrTriLeuIleuAlaGIYIleIleSerTriGIYGIuGIYCysAl 267
|||||
854 GACGGCGCTGCTGCTGCGCGGATCATCACTGGGGCGAGGGCTGTGC 903
267 aGIuArgAsnArgProGIYValTYrIleSerLeuSerAlaHisArgSerT 284
|||||
904 CGAGGGCAACAGGCCCGGGGTCTACATCAGCTCTGTGCGCACCGCTCT 953
284 rPValGIuIleValGIInGIYValGIInLeuArgGIYArgAlaGIInGIY 300
|||||
954 GGGTGAGAAAGATCGTCGACAGGGGTGCAGCTCCGCGGGCGCTCGAGGG 1003
301 GIYGIYAlaLeuArgAlaProSerGIInGIYSerGIYAlaIaIaArgSe 317
|||||
1004 GGTGGGGCCCTCAGGGCACCCAGCGGCTCTGTGGGCGCGCGGCTCTC 1053
317 r 317
1054 C 1054
seq_name: /SISL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AA\$1621
seq_documentation_block:
ID AA\$1621 standard; cDNA; 1352 BP.
XX
AC AA\$1621;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #837.
Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
ligase; hyperproliferative disorder; immunodeficiency disorder;
autoimmune disorder; neurological disorder; metabolic disorder;
inflammatory disorder; cardiovascular disorder; reproductive disorder;
blood-related disorder; infectious disorder; gene therapy; cytostatic;
anti arthritic; nephrotropic; anticoagulant; ss.
KM
KW
XX
OS Homo sapiens.
XX
PN WO20015301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.

201 ProIleIleAspSerGIuValCysSerHisLeuTyrTrpArgIValaGI 217
|||||
617 CCTATCATGACTCGAAGTCTGAGCATCTGTACTGGGGGAGACAG 666
217 YGIuGIYProIleThrGIuAspMetLeuCysAlaGIYTrpLeuGIuGIY 234
|||||
667 ACAGGACCATCATCTAGAGCATGCTGTGGCGGCTACTTGAGAGGG 716
234 IuArgAspAlaCysLeuGIYAspSerGIYProLeuMetCysGIuVal 250
|||||
717 AGCGGGAATGTTGTCTGGCGACATCCGGGGCCCTCATGTGCGAGTG 766
251 AspGIYAlaTrpLeuLeuAlaGIYIleSerTrpGIYGIuGIYCysAl 267
|||||
767 GAGGGCGCTGCTGCTGCTGGCGCATCATCGTGGGGGAGGGCTGTGC 816
267 aGIuArgAsnArgProGIYAlaTyrIleSerLeuSerAlaHisArgSerT 284
|||||
817 CGAGGCAACAGCGCGGGGTCTACATCATGCTCTGCGGACCGCTCCT 866
284 rValGIuIuTrIleValGIuGIuValGIuLeuArgGIYArgAlaGIY 300
|||||
867 GGCTGAGAGATCTGTGCAAGGGGTGACGCTCCGGGGCGGCTCAGGG 916
301 GIYGIYAlaLeuArgAlaProSerGIuGIYSerGIYAlaAlaArgSe 317
|||||
917 GGTGGGGCCCTCAGGACCGAGCGCTGTGGGGCGCGCGCTC 966
317 r 317
967 C 967
seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT: AAS26943
seq_documentation_block:
ID AAS26943 standard; CDNA: 1352 BP.
XX
AC AAS26943;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human CDNA encoding a novel secreted protein, SEQ ID 135.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;
cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
vulnary; secreted protein; rheumatoid arthritis;
hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
cerebrovascular disorder; cerebral ischemia; angiogenesis;
nervous system disorder; Alzheimer's disease; infection; ocular disorder;
corneal infection; wound healing; epithelial cell proliferation;
skin aging; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155441-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01320.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0226279.
PR 14-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.

|||||
567 AAGATGAGTTCCTTGCCCAACCCAGACCCCTGCAGAGCTGAAGTT 616
201 ProIleIleAspSerGluValCysSerHisLeuYrTrpArgGlyAlaG 217
617 CCTATCATGACTCGAGACTGTGCAGCCATCTGTACTGCGGGAGACAG 666
217 yGInGlyProIleThrGluAspMetLeuCysAlaGlyYrLeuGInGly 234
667 ACAGGAGACCCATCCTACTGAGGACATGCTGTGCGGCTACTTGAGAGG 716
234 luArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGInVal 250
717 AGCGGATGCTGTGCTGCGGCGACTCCGGGGGCCCTCATGTGCCAGGTG 766
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGInGlyCysAl 267
767 GACGGGCGCTGGCTGCTGGCCGCATCATCAGCTGGGGCGAGGGCTGTGC 816
267 agluArgAsnArgProGlyValYrIleSerLeuSerAlaHisArgSer 284
817 CGAGGCGCAACAGGCCCGGGGTCTACATCAGCCTCTGTGCGCACCGCTCT 866
284 rPValGluLysIleValGInGlyValGInLeuArgGlyArgAlaGInGly 300
867 GGTGGAGAGATCGTGCAGGGGTGCAGCTCCGGCGCGGCTCAGGGG 916
301 GInGlyAlaLeuArgAlaProSerGInGlySerGlyAlaAlaAlaArgSe 317
917 GGTGGGGCCCTCAGGGCAACCGAGCCAGGGGCTCTGGGGCGCGCGCGCTC 966
317 r 317
967 C 967

340 TGGGTGGTCAAGCCGCGACTGCTTTAAGAGCAATATGGACAACCATC 395

100 rLpHSeSerValIleuLeuGluAlaTrpGlnIleuGlnSnpProGlySera 117
|||||
396 TGTGTCTCAAGTATGTTGGGGGCGCTGGAGAGCTGGGGAGCCAGAGCCAA 445
|||||
117 TgSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSer 133
|||||
446 GGTCCCGAAGAAGTAGGCATGTCTGGGGTCTGCTCACCCAGGATATCT 495
|||||
134 TrpLysGlnGlyAlaCysAlaSplIleAlaLeuValAlaGluGluValGse 150
|||||
496 TGGAGAGAGGAAACCATGAGACATGCGCGTGGCGCCCTGGAAACCTC 545
|||||
150 rLleGlnPheSerGluLysValIleuProIleCysLeuProAspAlaSer 167
|||||
546 CATCATGTTCTGTAGCGGATCTCGCATCTGCTACTGACTACCTCTG 595
|||||
167 LeuHisLeuProProAsnTrpHisCysTrpIleSerGlyTyrGlySerIle 183
|||||
596 TCGGTCTCCCTCCCAAGACCGACTCTGGATTGCCGGCTGGGAGACATC 645
|||||
184 GlnAspGlyValProLeuProHisProGlnTrpLeuGlnLysLeuSva 200
|||||
646 CAGAGTGGAGTGGCCCTGCGCCACCTTAGACCTTAGAGAGCTGAAGGT 695
|||||
200 LProIleIleAspSerGluValCysSerHisLeuTyrTrpArgIValAg 217
|||||
696 GCCCATCATCTGACTCCGAATCTGCAAAGCTTGTACTGCGGGAGCGG 745
|||||
217 LysGlnGlyProLierhrGluAspMetLeuCysAlaGlyTyrTrpLeuGluGly 233
|||||
746 GTAGAGAACCATCACCGAGGCGCATGCTGTGTGGTACTCTGGAAGGg 795
|||||
234 GlnArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnVa 250
|||||
796 GACCGGAGATGCTTGTCTGGGCGACTCTGGGGGTCCCTGATGGCCAGGt 845
|||||
250 LaspGlyAlaTrpLeuAlaGlyIleIleSerTrpGlyGlnGlyCysa 267
|||||
846 GGATGACCATCTGCTACTGCTGCATTAACGTGGGGAGAGACGATGGG 895
|||||
267 IagLlnArgAsnArgProGlyValTyrTrpIleSerLeuSeraHisArgSer 283
|||||
896 CGAGAGCGAACCGGCCGGTGTGACACACAGCGCTCCAGCTACCGCTCC 945
|||||
284 TrpValGlnLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnI 300
|||||
946 TGGGTGCAAAAGATCTGTCAAGGGGTCAAGCTCGCGGTACTTGGCGGA 995
|||||
300 yGly.GlyAlaLeuArgAlaProSerGlnGlySer 311
|||
996 CAGTGGGGACACAGAAAGCTCTTAATCTAGAGATCT 1030
|||
seq_name: qb_esc12:BE907706
seq_documentation_block:
LOCUS BE907706 470 bp mRNA linear EST 20-OCT-2000
DEFINITION 601497374FL NIH_MGC_70 Homo sapiens cdna clone IMAGE:3899480 5',
mRNA sequence.
ACCESSION BE907706
VERSION BE907706.1 GI:10401537
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.


```

67  eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArgT 84
    |||||
255  CCAAGAAATGAGGACCCACCATGCGCAGAGTCTGCTGCACAGCCGGT 304
    |||||
84  rPvalIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
    |||||
305  GGGGTGACACTGCTGCCACTGTTTCAGAGCAACCTGAACAAACCATAC 354
    |||||
101  LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
    |||||
355  CTGTCTCTGTGCTGCTGGGGGCTGCGACACTGGGAACCTGCTCTCG 404
    |||||
117  gSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSerT 134
    |||||
405  GTCCAGAAAGTGGTGTGCTGCTGGGTGAGCCGCCCTGTGTAT.... 450
    |||||
134  rPySgInGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
    |||||
450  .....
51  IleGlnPheSerGlnArgValLeuProIleCysLeuProAspAlaSerTl 167
    |||||
450  .....
167  eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerTleG 184
    |||||
450  .....
184  LuAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuLysVal 200
    |||||
450  .....
201  ProIleIleAspSerGlnValCysSerHisLeuTyrTrpArgGlyAlaG 217
    |||||
450  .....
217  yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGluGlyG 234
    |||||
450  .....
234  LuArgAspAlaCysLeuGlnGlyAspSerGlyGlyProLeuMetCysGlnVal 250
    |||||
450  .....
251  AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAl 267
    |||||
450  .....
267  aGluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSerT 284
    |||||
451  .....TCTT 454
    |||||
284  rPvalGlnLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
    |||||
455  GGGTGGAGAAAGATCGTGCAGAGGGTGCAGCTCCGGGGCGGCTCAGGG 504
    |||||
301  GLyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
    |||||
505  GGTGGGGCCCTCAGGGGACGAGCCAGGGCTCTGGGGCCGCCGCCGCTC 554
    |||||
317  r 317
    |||||
555  C 555

```

seq_name: gb_esl1:AW462236

seq_documentation_block:

LOCUS AW462236 473 bp mRNA linear EST 24-FEB-2000

DEFINITION BP230008B20C12 Soares normalized bovine placenta Bos taurus cDNA

ACCESSION AW462236

VERSION AW462236.1 GI:7032404

```

KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euteria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
1 (bases 1 to 473)
Lewin,H.A., Soares,M.B., Rebeltz,M., Pardinas,J., Liu,L. and Larson
J.H.
TITLE Bovine ESTs
JOURNAL Unpublished (2000)
COMMENT Contact: Lewin, H. A.
W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Funding for cattle EST sequencing was provided by the USDA National
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
from Washington University Genome Center. Vector Trimmi g:
Cross_match from Washington University Genome Center PHRAP
Sequences submitted are vector free and at least 200 bp in length.
PCR primers
FORWARD: TAATGACTCAGTATAGG
BACKWARD: APTAACCTCAGTAAG
Insert Length: 473 Std Error: 0.00
Plate: BP230008B20 row: C column: 12
Seq primer: ACGGATTAACAATTCACACAGA
High quality sequence stop: 473.
Location/Qualifiers
1..473
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone="BP230008B20C12"
/clone_11b="Soares normalized bovine placenta"
/sex="female"
/lab_host="DH10B"
/note="Organ: Placenta; Vector: pRTTpac; Site:1; Ecoli;
Site:2; Note: The cDNA library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. "
BASE COUNT 79 a 178 c 138 g 78 t
ORIGIN
alignment_scores:
quality: 732.00 length: 157
ratio: 4.913 gaps: 0
Percent Similarity: 94.904 Percent Identity: 82.166
alignment_block:
US-10-040-803-7 x AW462236 ..
Align seq 1/1 to: AW462236 from: 1 to: 473
66 SerIleGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSer 82
|||||
3 AGCATCCGGAAGAACGACCCACCACTGCGGGCTCCCTGCTCAGCAG 52
|||||
82 rArgTrpValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysP 99
|||||
53 CCGCTGGGTGCTCAGCGCCGCCACATGCTTCAGAGTAATTCAGACAAC 102
|||||
99 rGlyTrpLeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlnLysProGly 115
|||||
103 CAACCCAGTCTCTGTGCTGCTGGAGGCTGCGCAAGCTGGGAAACCTCGGC 152
|||||
116 SerArgSerGlnLysValGlyValAlaTrpValGluProHisProValTy 132
|||||

```



```

153 CCAAGGTCCAGAGGTGGTATGCGCTGGGACAGCCCACTCTGTGA 202
132 rsertrpysgluglyAlaCysAlaAspIleAlaLeuValArgLeuIua 149
|||||
203 CTCTGGAGAGGGCTCCCGGACATCGCTGCTGGCTGGAGC 252
149 rSerIleInPheSerGluArgValLeuProIleCysLeuProAspAla 165
|||||
253 GCGCATTCAGTCTCTGAGCGCGCTCGCCATCTGCTGCCGAGCTCC 302
166 SerIleHisLeuProAsnThrHisCysTrpIleSerGlyTrpGlyse 182
|||||
303 ACCGTCCACTCTCTCCGACACCAACTGTGATTGCCGGCTGGAGAG 352
182 rIleGlnAspGlyValProLeuProHisProGlnIleThrLeuGlnLeu 199
|||||
353 CGTCCAGATGAGAGTCCCTCCCTCCAGCTGAGACCTCCAGAGCTGA 402
199 yValProIleIleAspSerGluValCysSerHisLeuTrpArgly 215
|||||
403 AGGTCCCATCATCGACTGACGACCTGAGCGCGCTGTACTGCGGGGA 452
216 AlaGlyGlnGlyProIleThr 222
|||||
453 GCGGCGAGGGCGCATCAC 473

```

seq_name: gb_est2:BF180373

seq_documentation_block:

LOCUS BF180373 935 bp mRNA linear EST 31-OCT-2000
DEFINITION 601804557F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035351 5',
mRNA sequence.
ACCESSION BF180373
VERSION BF180373.1 GI:11058515
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 935)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-riemail.nih.gov
Tissue Procurement: Lotmar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM9309 row: a column: 16
High quality sequence stop: 732.
Location/Qualifiers
1..935

FEATURES

source

/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:4035351"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SpOrf6; Site:1: salI;
Site:2: NotI; Cloned unidirectionally. Primer: oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lotmar Hennighausen/Robin Humphreys,
NIH"

BASE COUNT 197 a 279 c 287 g 172 t
ORIGIN

alignment_scores:
Quality: 710.50 Length: 310
Ratio: 3.230 Gaps: 9
Percent Similarity: 70.968 Percent Identity: 55.806

alignment_block:
US-10-040-803-7 x BF180373

Align seg 1/1 to: BF180373 from: 1 to: 935

```

1 MetValAlaSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyTh 17
|||||
81 ATGATGATTCAGACACCTCCACACACTGGGTGGGACAGATTACAGAT 130
|||||
17 rPheThrSerLeuLeuLeuAlaSerThAlaIleLeuAsnAlaAla 34
|||||
131 CTAAATCCTTGTGTGCTGCTGACTTCCACAGCTCCATCAGTGTGCA 180
|||||
34 rGleProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
|||||
181 CCATCCAGATGTCGCCAGACTGTGGAAAGCTCAGCAGCTGAAACCGATT 230
|||||
51 ValGlyGlyGluAspSerThrAspSerGluTrpProTrpIleValSerI 67
|||||
231 GTGGAGGTGAGGACAGCATGATGCCAGTGGCCCTGGATTGTAGCAT 280
|||||
67 eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArg 84
|||||
281 CTCACAGAAATGGCTCCACACACTGTGACAGCTCCCTGCTCACCAACCT 330
|||||
84 rValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
|||||
331 GGGTGTGCACAGCGCGCACTGCTTAAAGACATATGACCAACACATCT 380
|||||
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
|||||
381 CTGTTCACAGTATTGTTGGGGCTGGAGCTGGGAGCCAGCCCAAG 430
|||||
117 gSerGlnLysValGlyAlaAlaTrpValGlnProHisProValIleTyrSer 134
|||||
431 GTCCCAAAAGTAGGCAATTCCTGGGTGCT.GCTCACCCAGCTATCTT 479
|||||
134 rPylGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSer 150
|||||
480 GGAAGGAGGAGACCATGCGACATGTCCTGCTGCGCTGGAAACATGCC 529
|||||
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167
|||||
530 ATCCAGTTCCTGAGCGGATCTGCCCATCTGCTA.CCTGACTCTCTGT 578
|||||
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerIle 184
|||||
579 CCGTCTCCCTCCCAAGACGAC.TGCTGATGTGGCG.TGGGGAAGGATCC 626
|||||
184 IlnAspGlyValProLeuProHisProGlnIleThrLeuGlnLysLeuVal 200
|||||
627 AGGATGGGGTCCCGCGGCC...ACCTCAGACCTTCAAGAGCGAAGTGG 673
|||||
201 ProIleIleAspSerGluArgValCysSerHisLeuTyrTrpArgGlyAla 217
|||||
674 CCA...TCCTGACTCGAAGCTGGAAGAGCTGTAC...CGGGGAGAGCG 717
|||||
217 yGlnGlyProIleThrGlnAspMetLeuCysAlaGlyTyrLeuGlnGlyG 734
|||||
718 GTCAGGAGCCATCACGGGCT.....GCTGATGCTGCTGACTGGAG 758
|||||
224 IuArgAspAlaCysLeuGlyAspSerGlyLysProLeuMetCysGlnVal 250
|||||
759 GAAGGAAAGCGGTGTGGGACT....GGGTCGGAGATGTCCGGGTG 802
|||||
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyLeuGly..... 265
|||||
803 .....GCGGCTGCGGAGAACCGAGGAGGGGCGGAG 834

```


High quality sequence stop: 507.

FEATURES
Location/Qualifiers
1..507

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="H0345"
/dev_stage="Adult"
/note="Organ: head/neck; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 90 a 171 c 129 g 117 t
ORIGIN

Percent scores:

Quality: 657.00 Length: 126
Ratio: 5.256 Gaps: 0
Percent Similarity: 99.206 Percent Identity: 97.619

alignment_block:
US-10-040-803-7 x AM383315 ..

Align seg 1/1 to: AM383315 from: 1 to: 507

```

95 AsnLeuAsnLysProTyrLeuPheSerValLeuLeuGlyAlaTyrGlnLe 111
|||||
128 AACCTGAACAACCAACCTGTTCTGTGCTGCTGGGAGGC.TGCCACCT 176
|||||
111 uGlyAsnProGlySerArgSerGlnLysValGlyAlaTyrPValGluP 128
|||||
177 GGGGAACCTGGCTCTGCTGCTCCAGAGGTGGGTCTGCTGGTGGAGC 226
|||||
128 roHisProValTyrSerTyrPlyGlnGlyAlaCysAlaAspIleAlaLeu 144
|||||
227 CCCACCTGTGTATTCCTGGAAAGAGGGCTGTGCAGACATTCCTGCTG 276
|||||
145 ValArgLeuGluArgSerIleGlnPheSerGluArgValLeuProIleCy 161
|||||
277 GTGCGTCTGAGCGCTCCATACAGTTCTCAGACGGGTCTGCCCATCTG 326
|||||
161 sleuProAspAlaSerIleHisLeuProAsnThrHisCysTrpIle 178
|||||
327 CCTACTGATGCTCTATCCACCTCCCAACACCCACTGCTGGATCT 376
|||||
178 erGlyTyrPlySerIleGlnAspGlyValProLeuProHisProGlnThr 194
|||||
377 CAGGCTGGGGGAGCAGCCAAAGATGAGTTCCTGCCACCCCTCAGACC 426
|||||
195 leuGlnLysLeuLysValProIleIleAspSerGluValLysSerHisLe 211
|||||
427 CTGCAGAAAGCTGAAGGTTCTATCATCATCAGTCGGAAGTCTCAGACCA 476
|||||
211 uTyrTrpArgGlyAlaGlyGlnGlyPro 220
|||||
477 GTACTGGGGGAGATGAAACAGAGGAGCC 504

```

seq_name: gb_est1:A1940071

seq_documentation_block:

LOCUS A1940071 330 bp mRNA linear EST 03-AUG-1999
DEFINITION IL2-CT0031-290799-001-A12 CT0031 Homo sapiens cDNA, mRNA sequence.
ACCESSION A1940071
VERSION A1940071.1 GI:5687052
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 330)

AUTHORS
TITLE
JOURNAL
COMMENT

HCGP <http://www.ludwig.org.br/ORESTES>.
The FAPESP/LICR Human Cancer Genome Project
Unpublished (1999)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?rl=IL2&rl=IL2-CT0031-290799-001-A12&rl=1999-07-29&rl=1>)
Seq primer: puc 18 forward
High quality sequence stop: 330.

FEATURES

source

1..330
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0031"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 69 a 101 c 101 g 59 t
ORIGIN

alignment_scores:

Quality: 600.00 Length: 109
Ratio: 5.505 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-10-040-803-7 x A1940071/rev ..

Align seg 1/1 to reverse of: A1940071 from: 1 to: 330

```

53 GlyLysPserThrAspSerGluTrpProTyrIleValSerIleGlnLy 69
|||||
329 GGCAGGAGACACACCTGACAGCGAGTGGCCCTGGATCTGAGCATCCAGAA 280
|||||
69 sAsnGlyThrHisCysAlaGlySerLeuLeuThrSerArgTrpValI 86
|||||
279 GAATGGAGACCCACACCTGCGCAGGTTCTGTCTCACCAGCCGCTGGGTGA 230
|||||
86 leThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyrLeuPhe 102
|||||
229 TCACTGCTGCCCACTGTTTCAAGACAACTGAACAACATCATCTGTTTC 180
|||||
103 serValLeuLeuGlyAlaTyrPlyGlnLeuGlyAsnProGlySerArg 119
|||||
179 TCTGTGCTGTGGGGGCTGTGAGCTGGGGAGACCTGGCTCTGCTGCCCA 130
|||||
119 nLysValGlyValAlaTyrPValGluProHisProValTyrSerTrpLys 136
|||||
129 GAAGGTGGGTGTGGCTGGGTGGAGGCCCACTGTGTATCTCTGGAAAG 80
|||||
136 lucGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSerIleGln 152
|||||
79 AAGGTGCTGTGCAGACATTTGCCCTGTGCTCGACGCTCCATACAG 30
|||||
153 PheSerGluArgValLeuProIleCys 161
|||||
29 TTCTCAGAGCGGGTCTCGCCCATCTGC 3

```

seq_name: gb_est2:BM389391

seq_documentation_block:
 LOCUS BM389391 703 bp mRNA linear EST 17-JAN-2002
 DEFINITION UI-R-CNI-cj-k-1-03-0-UI.s1 UI-R-CNI Rattus norvegicus cDNA clone
 UI-R-CNI-cj-k-1-03-0-UI 3', mRNA sequence.
 ACCESION BM389391
 VERSION BM389391.1 GI:18189444
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus
 1 (bases 1 to 703)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 CONTACT: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized cervix library cDNA Library Preparation: M.B. Soares Lab
 Clone distribution: clones will be available through Research
 Genetics (www.resgen.com) The following repetitive elements were
 found in this cDNA sequence: 1-36, >AT-richLow-complexity
 Seg primer: M13 Forward
 POLYA=Yes.

FEATURES
 source Location/Qualifiers
 1..703
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-CNI-cj-k-1-03-0-UI"
 /clone_1lb="UI-R-CNI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="vector: p7773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CNI
 library is a subtracted library derived from the following
 pool of seven normalized rat libraries: normalized rat
 seminal vesicles, normalized rat penis, normalized rat
 bladder, normalized rat cervix, normalized rat brown
 adipose, normalized rat fundus, and normalized rat
 salivary gland. It was constructed according to the
 procedure described by Bonaldo, Lennon & Soares (Genome
 Research Genome 6: 791-806, 1996). For construction of
 the CNI library, plasmid DNA from the pool of normalized
 libraries was electroporated into competent bacteria for
 the production of single-stranded circular DNA. This was
 then used as a tracer in a subtractive hybridization with
 a driver (PCR amplified inserts from a plasmid DNA template
 preparation) comprising: a) a pool of about 34,000 clones
 from the Rat Unigene Set corresponding to plates R-5-AA-NN
 excluding plates R-5-MM and MN. This pool represented 40%
 of the final driver population. b) a pool of about 29,000
 clones from subtracted libraries CNO and CAL corresponding
 to plates R-CA0-AMV through R-CA0-AXS, R-CA0-AXZ through
 R-CA0-BAL, R-CA0-BFE through R-CA0-BHJ, R-CA0-BOS,
 R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-K, R-CA0-BKP through
 R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BMA,
 R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through
 R-CA0-BOI, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through
 R-CA1-BDA, R-CA1-BHZ through R-CA1-BJF, R-CA1-BBR,

R-CA1-BJT through R-CA1-BKB, R-CA1-BKD, R-CA1-BKF,
 R-CA1-BKI, R-CA1-BKT, R-CA1-BLV, R-CA1-BLH through
 R-CA1-BLN, R-CA1-BLS, R-CA1-BLU-V, R-CA1-BNR, and
 R-CA1-BLE. The resulting pool represented 20% of the
 final driver population. c) a pool of about 15,000 clones
 from non-normalized libraries CS0, CT0, CU0, CW0, and CX0
 and normalized libraries CS0, CT0, CU0, CW0, and CX0
 corresponding to plates R-CS0s-CBD through R-CS0s-CBO,
 R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through
 R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN
 through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV
 through R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN,
 R-CU0-BUQ through R-CU0-BVL, R-CW0-BVV through R-CW0-BMP,
 R-CW0-BXN through R-CW0-BXO, R-CX0-BMQ through R-CX0-BXM.
 The resulting pool represented 5% of the final driver
 population. d) a pool of about 5,000 clones (1,000 from
 non-normalized eye library CV0 and 4,000 from normalized
 eye library CV1) corresponding to plates R-CV0-BRH through
 R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through
 R-CV1-BTC, and R-CV1-BVO through R-CV1-BVO. This pool
 represented about 5% of the final driver population. e) A
 pool of about 10,000 clones from subtracted library BS2,
 BV0 and BV0P (7-9.5 kb cDNA library fraction from rat
 whole embryo), and BX0 (0.5-7kb cDNA library fraction from
 rat whole embryo) corresponding to plates R-BS2-BDB
 through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0P-AOI
 through R-BV0P-AOX, and R-BX0-AGY through R-BX0-ASH. The
 resulting pool represented 5% of the final driver
 population. f) a pool of about 7,000 clones from the
 seven non-normalized libraries that make up the tracer
 including CV0, CZ0, DAO, DB0, DCO, DD0, and DE0
 corresponding to plates R-CY0-BXP through R-CT0-BXZ,
 R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BZB-C, R-DAO-BYJ
 through R-DAO-BYP, R-DAO-BZD through R-DAO-BZH, R-DB0-BYQ
 through R-DB0-BZA, R-DC0-BZI through R-DC0-BZO, R-DC0-CAY
 through R-DC0-CBA, R-DD0-BZR through R-DD0-CAL,
 R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The
 resulting pool represented about 10% of the final driver
 population. g) a pool of about 2,000 clones from the pool
 of normalized libraries, CNO, that makes up the tracer.
 The corresponding plates are R-CNO-BKW through R-CNO-BLD,
 R-CNO-BLG, R-CNO-BLP through R-CNO-BLP, R-CNO-BLF,
 R-CNO-BLM-X, R-CNO-BMB, and R-CNO-BMF through R-CNO-BML.
 This pool represented 5% of the final driver population.
 h) a pool of the 28 most abundant clones in the CNO pool
 corresponding to the following addresses: bkx-a-09-0-UI,
 bkx-b-09-0-UI, bkx-b-11-0-UI, bkx-b-10-0-UI, bkx-d-01-0-UI,
 bkx-d-06-0-UI, bkx-g-08-0-UI, bkx-h-12-0-UI,
 bkx-a-05-0-UI, bkx-a-06-0-UI, bkx-a-11-0-UI, bkz-c-06-0-UI,
 bkz-c-09-0-UI, bkz-d-10-0-UI, bla-a-01-0-UI,
 bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI,
 blb-e-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI,
 blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blg-h-04-0-UI
 , blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5%
 of the final driver population. i) One abundant CNO clone
 (corresponding to the address bkx-a-11-0-UI) was digested
 with Not I and Eco RI and the resulting insert was gel
 purified. This purified insert was added directly to the
 driver so that it represented 5% of the final driver
 population.
 TAG_LIB=UI-R-CNI
 TAG_TISSUE=cervix
 TAG_SEQ=GACCA*

BASE COUNT 168 a 201 c 178 g 155 t 1 others
 ORIGIN

alignment_scores:
 Quality: 586.00 Length: 133
 Ratio: 4.883 Gaps: 0
 Percent Similarity: 90.226 Percent Identity: 82.707
 alignment_block:

US-10-040-803-7 x BM389391/rev ..

Align seg 1/1 to reverse of: BM389391 from: 1 to: 703

```

170 ProProAsnThrHisCysTrpIleSerGlyTrpGlySerIleGlnAspG1 186
|||||.....:|||||.....:|||||.....:|||||.....:|||||
702 CCTCCACACCAACTGCTGATGTCGCCGCTGGGAGAACATCCAGAGATG 653
186 yValProLeuProHisProGlnThrLeuGlnLysLeuLysValProIle1 203
|||||.....:|||||.....:|||||.....:|||||.....:|||||
652 AGTGCCCTGCACCCCTCAGACCCCTCAGACCTCAGACCTCAGATCA 604
203 leAspSerGluValCysSerHisLeuTrpTrpArgGlyAlaGlyGlnGly 219
|||||.....:|||||.....:|||||.....:|||||.....:|||||
603 TCGACCCCTGAACCTGCAAAAGTTGTACTGCGCGGAGCTGGTCAGGAA 554
220 ProIleThrGluAspMetLeuCysAlaGlyTrpLeuGlnGlyGluArgAs 236
|||||.....:|||||.....:|||||.....:|||||.....:|||||
553 GCCATCACCGAGGACATCTGTCGTCTGTACTCGAAGGAGAGCGGGA 504
236 PALAcysLeuGlyAspSerGlyGlyProLeuMetCysGlnValAspGlyA 253
|||||.....:|||||.....:|||||.....:|||||.....:|||||
503 CGCTGTGTGGGCGACTGTGGGGGTCGCCCTGATGTGCCAGGTGATGACC 454
253 latrPLeuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAlaGluArg 269
|||||.....:|||||.....:|||||.....:|||||.....:|||||
453 ACTGGCTTACTGACGGGCAATATCACTGGGAGAGAGGCTGCGCGAGGCG 404
270 AsnArgProGlyValTyrIleSerLeuSerAlaHisArgSerTrpValG1 286
|||||.....:|||||.....:|||||.....:|||||.....:|||||
403 AACCGCGCGGCGGTACACACGACTCTACCTCACCCTCCCTGGGTGCA 354
286 ulysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGlyGly 302
|||||.....:|||||.....:|||||.....:|||||.....:|||||
353 GAGGATCTGTGCAAGGGGTGCAAGCTGCGAGGCGCTTGGCGGACAGTGG 305

```

seg_name: qb_est2:BG480197

seg_documentation_block:

```

LOCUS BG480197 1001 bp mRNA linear EST 21-MAR-2001
DEFINITION 602530380F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:465395 5',
mRNA sequence.
ACCESSION BG480197
VERSION BG480197.1 GI:13412476
KEYWORDS EST.
SOURCE human.
ANALYSIS Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1001)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph. D.
COMMENT Email: c9apbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
CDNA Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1440 row: P column: 20
High quality sequence stop: 536.
Location/Qualifiers
1..1001
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:465395"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; site_1: XhoI;

```

Site 2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(C). Size selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```

BASE COUNT 254 a 255 c 292 g 200 t
ORIGIN

```

```

alignment_scores:
Quality: 580.50 Length: 267
Ratio: 2.977 Gaps: 24
Percent Similarity: 73.034 Percent Identity: 64.045

```

alignment_block:

US-10-040-803-7 x BG480197 ..

Align seg 1/1 to: BG480197 from: 1 to: 1001

```

1 MetValValSerGlyAlaProProAla.LeuGlyGlyGlyCysLeu.Gly 16
|||||.....:|||||.....:|||||.....:|||||.....:|||||
53 ATGGTGGTTTCTGGAGCGCCACCCAGACCTGGGTGGGCTGTCTGGC 102
17 ThrPheThrSerLeuLeuLeuAla.SerThrAlaIleLeuAsnAla.. 32
|||||.....:|||||.....:|||||.....:|||||.....:|||||
103 ACCCTTACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 152
33 AlaArgIleProValProProAlaCysGlyLysProGlnGlnLeuAsnAr 49
|||||.....:|||||.....:|||||.....:|||||.....:|||||
153 GCCAGATACCTGTTCCGCCAGCCTGTGGGAGAACCCAGACCTGAACCG 202
49 gValValGlyGly.GluAspSerThrAspSerGluTrpProTrpIleVal 65
|||||.....:|||||.....:|||||.....:|||||.....:|||||
203 GGTGTGGGGCGGTGAGGACAGACTGACAGCGAGTGAGCTGTGATCTG 252
66 SerIleGlnLysAsn.GlyThrHisCysAla.GlySerLeu.LeuTh 81
|||||.....:|||||.....:|||||.....:|||||.....:|||||
253 AGCATCCAGAGAGATAGGAGCCACCCAGCTGCGAGATTTCTAGCTCAC 302
81 rSerArgTrpValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnL 98
|||||.....:|||||.....:|||||.....:|||||.....:|||||
303 CAGACGCTGGGTGATTCACCTGCTGCCCTGTTCAAGAGCAACCTGAACA 352
98 ySProTyrLeu.PheSerValIleLeuGlyAlaTrp...GlnLeuGlyAs 113
|||||.....:|||||.....:|||||.....:|||||.....:|||||
353 AAACATACCTGTTACTCTGTGCTGCTAGGGGCTGGCAAGCTGGGGA 402
113 nProGly...SerArgSerGlnLysValGlyValAlaTrp.ValGlu.Pr 128
|||||.....:|||||.....:|||||.....:|||||.....:|||||
403 CCCATGCTCTCGAGTCCCAAGAGAGGTGGTGTGCTGGGATGAGAGACC 452
128 OHAspProValTyrSerTrpLysGlnGlyAlaCysAlaAsp...IleAla 144
|||||.....:|||||.....:|||||.....:|||||.....:|||||
453 CCACCTGTGTATTCCTGAGCGAGAGAGTGGCTGTGACAGCAATGTCTCG 502
144 euValArgLeuGluArgSerIleGln.PheSerGlu.ArgValLeuProI 160
|||||.....:|||||.....:|||||.....:|||||.....:|||||
503 GATCAGCTGTGAGCGCTCCATCAAGTTCTCAGAGCGGGGTGTGCA 552
160 leCysLeuProAspAla.....SerIleHisLeuProProAsnThrHis 174
|||||.....:|||||.....:|||||.....:|||||.....:|||||
553 TCTGGCTACTGATGCTCTTATCCAACTCAACACGTACAGAGAACGCAAT 602
175 CysTrpIleSerGlyTrp...GlySerIleGlnAspGly.ValProLeu. 189
|||||.....:|||||.....:|||||.....:|||||.....:|||||
603 GAGTGAGATCTCAGAGCTAGAGGAGGCCATCCAAAGATGAATCACTTG 652
190 ..ProHisProGlnThrLeuGlnLysLeu...LysValPro.IleIleAs 204
|||||.....:|||||.....:|||||.....:|||||.....:|||||
653 CACCCAACTCAAGAGCTGAGAGAAAGCTGAGAAAGTTCTTACTATCGA 702

```

```
204 pSerGluValCysSerHisLeuTyrTrpArgGlyAlaGly.GlnGlyPro 220
    |||  ::  ::  ::  ||  ::  ||| ||| |||
703 CTCGGCAGTCCGACAGACATCTGTACTTGGCGGAGGACAGACAGACCA 752
    |||  ::  ::  ||  ::  ||| ||| |||
221 IleHrGluAspMet.....LeuCysAlaGlyTyrLeuGluGlyG1 234
    |||  ::  ::  ||  ::  ||| ||| |||
753 GGGACCCAGTTCATCTGAGACACTGCTGTGTGTGACCGAGATACTGGAAA 802
    |||  ::  ::  ||  ::  ||| ||| |||
234 uArgAspAlaCysLeuGly 240
    :  |||  ::  ::  ||  ::  ||| ||| |||
803 AGGTAACGACGAGATTGGA 821
```

53 IYGLuaspSerThraspSerGluTrpProTrpI

```

/prodata/1/ina/5A.COMB.seq:US-08-978-404B-19 + 449.50
/prodata/1/ina/5A.COMB.seq:US-08-920-900A-1 + 443.50
/prodata/1/ina/PCPUS.COMB.seq:PCP-US94-00616-1 + 443.50
/prodata/1/ina/5B.COMB.seq:US-09-027-337-1 + 438.50
/prodata/1/ina/5A.COMB.seq:US-08-508-448C-15 + 434.50

ne: /cgn2_6/prodata/1/ina/5B.COMB.seq:US-09-016-366A-14

mentation block:
umentation 14, Application US/09016366A
t No. 5955431
RAL INFORMATION:
PLICANT: Stevens, Richard L.
PLICANT: Huang, Chifu
TLE OF INVENTION: M5T CELL PROTEASE PEPTIDE
TLE OF INVENTION: INHIBITORS
MBER OF SEQUENCES: 65
DRESSEE: Wolf, Greenfield & Sacks, P.C.
REET: 600 Atlantic Avenue
ity: Boston
STATE: MA
OUNTRY: U.S.A.
ZIP: 02210-2211
OMPUTER READABLE FORM:
EDIUM TYPE: Diskette
OMPUTER: IBM Compatible
PERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
URRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,366A
ILING DATE: January 30, 1998
CLASSIFICATION: 530
RIOR APPLICATION DATA:
APPLICATION NUMBER: 60/037,090
ILING DATE: 05-FEB-1997
TTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7093
ELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
FORMATION FOR SEQ ID NO: 14:
EQUENCE CHARACTERISTICS:
LENGTH: 1108 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
016-366A-14

ent_scores:
Quality: 554.00 Length: 279
Ratio: 2.916 Gaps: 9
ent Similarity: 68.100 Percent Identity: 41.577

ent block:
-040-803-7 x US-09-016-366A-14 ..

a seg 1/1 to: US-09-016-366A-14 from: 1 to: 1108

21 leuLeuLeuLeuAaSerThrAlaIleuAsnAlaIahArgIleProVa 37
|||||
:::
49 CTCGTGCTGCTGGGACACTCTCCCTCGGTAGTCTGGGTACTACAC 98
|||||
37 lPro...ProAlaCysGlyIysProGInGInLeuAsnArgValaIaGly 53
|||||
99 CCTCGCGCCAGCC.....ATCACGAGGTGGCATCTGGGAG 133
lyGluAspSerThrAspSerGluTrpProTyrIleValSerIleGln... 68

```

```

137 GACATGAGGCTTCTGAGAGTAAGTGGCCCTGCGACGAGTGGAGCTGATTT 186
69 .....LysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerAr 83
187 AATATTAACTACTGATATCTTCTGCGAGGCTCTCTCATCCACCACA 236
83 gTTPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProT 100
237 GTGGGTGCTGCTGCGGACACACTGTGGGACCGCCACATATAAAGCCAC 286
100 yLeuPheSerValLeuLeuGlyAlaTrpLeu.....GlyAsnPro 114
287 AGCTCTTCGCGGTGACGTTCTGAGCAGTATCTATGAGGAGC... 333
115 GlySerArgSerGlnLysValGlyAlaTrpValGluTrpHisProVa 131
334 .....CAGCTCTCTCTTTTAAACCGGATGCTGGTGCACCCCA 371
131 lTySerTrpLysGluGlyAlaCysAlaAspIleAlaLeuValArgLeu 148
372 CTATTACACGCGCGAGGTTGG...GCAGACGTGGCCCTGCTGAGCTTG 418
148 lueSerIleGlnPheSerGlnArgValLeuProIleCysLeuProasp 164
419 AGGTCCCTGTAATGTCCTCCACCATTATCCATATCCCTGCCCCCT 468
165 AlaSerIleHisLeuProAsnThrHisCysTrpIleSerGlyTrpAl 181
469 GCCTCGGAGACTTCCCTCCGAGATGCTGCTGCTGACAGCGGGG 518
181 ySerIleGlnAspGlyValProLeuProHisProGlnThrLeuGlnLys 198
519 CGACATTGTAATGACGAGCTCTCCACTCTTATCTCTGAAAGCAG 568
198 euLysValProIleIleAspSerGlnValCysSerHisLeuTrpArg 214
569 TGAAGTTCCCATTTGGAAACAGCTCTGTGACCGGAAGTACCACT 618
215 GlyAlaGlyGlnGly.....ProIleThrGluAsp...MetLeuCy 227
619 GGCCTCTACACGAGATGATTTTCCCATTTGTCATGATGATGCTGTG 668
227 saLagIyTrpLeuGluGlyGluArgAspAlaCysLeuGlyAspSerGly 244
669 TGTCTGA.....AATACAGAGAGACTCTGCGACGCGCATTCAGGGG 712
244 lYProLeuMetCysGlnValaAspGlyAlaTrpLeuLeuAlaGlyIle 260
713 GGCACCTGCTGCAAGTGAAGGATGCTGCTGACGACGAGAGTGTGTC 762
261 SerTPGlyGluGlyCysAlaGluArgAsnArgProGlyValTrpIle 277
763 ACCTGGGGGAGGCTGCGCACAGCCCAACAGCTGCGCATCTACACCG 812
277 rLeuSerAlaHisArgSerTrpValGluLysIleVal 289
813 GGTGACATACCTTACAGCTGATCCACGCTATGTC 849
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-978-404B-20
seq_documentation_block:
; Sequence 20, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston

```

```

STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1108 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-978-404B-20

alignment_scores:
Quality: 554.00 Length: 279
Ratio: 2.916 Gaps: 9
Percent Similarity: 68.100 Percent Identity: 41.577

alignment_block:
US-10-040-803-7 x US-08-978-404B-20 ..
Align seg 1/1 to: US-08-978-404B-20 from: 1 to: 1108

21 LeuLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAlaArgIleProVa 37
||||| : : : : :
49 CTGCTGCTGCTGCGACACTGCTCCCTCGGCTAGTCTGCTGTAAGTACG 98
37 lPro...ProAlaCysGlyLysProGlnGlnLeuAsnArgValaGly 53
||| : : : : :
99 CCTGCGCCAGCC.....AATCAGCAGAGTGGCATGCTGGGAG 136
53 lYgluAspSerThrAspSerGluTrpProThrIleValSerIleGln... 68
||||| : : : : :
137 GACATGAGGCTTCTGAGAGTAAGTGGCCCTGCGACGAGTGAACCTGATTT 186
69 .....LysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerAr 83
||| : : : : :
187 AATTAATACACTGATGATATTTCTGCGGAGCTCTCTCATCCACCACA 236
83 gTTPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProT 100
: : : : :
237 GTGGGTGCTGCTGCGGACACACTGTGTGGGACCGCCACATATAAAGCCAC 286
100 yLeuPheSerValLeuLeuGlyAlaTrpLeu.....GlyAsnPro 114
||||| : : : : :
287 AGCTCTTCGCGGTGACGTTCTGAGCAGATATCTATGAGGAGC... 333
115 GlySerArgSerGlnLysValGlyAlaTrpValGluTrpHisProVa 131
||| : : : : :
334 .....CAGCTCTCTCTTTGAAACCGGATGCTGGTGCACCCCA 371
131 lTySerTrpLysGluGlyAlaCysAlaAspIleAlaLeuValArgLeu 148
||| : : : : :
372 CTATTACAGCGCGAGGTTGG...GCAGACGTGGCCCTGCTGAGCTTG 418

```


TELEFAX: 617-7

578 AGCCTGTGTGATCGGAMGTACCCACACATGCCCCCTACACACGGGAAAGTATG

220 .ProIleThrGluasp...MetLeuCysAlaGlyTyrLeuGluGlyGluA 2

Page 4

```

US-10-040-803-7 x US-09-016-366a-24 ..
Align seg 1/1 to: US-09-016-366a-24 from: 1 to: 1103

21 LeuLeuLeuLeuAlaSerThr.....AlaIleLeuAsnAlaIleAla 34
      |||||.....
29 CTCTCTGCTGTGGACACGTCCGCCCTGGCTAGCTGGTGACGGGCC 78
      |||||.....
34 gileProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 51
      |||||...||
79 TTGCCACGTAAACACAGACGTGGGC.....ATTG 107
      |||||.....
51 alAlgylgIyIuAspSerThrAspSerGluIrrProTrrpIleValSerIle 67
      |||||.....
108 TGGGAGACACAGAGGCTTCTGAAAGTAAGAGGCCCTGGCAGGTAGGCTG 157
      |||||.....
68 GlnLysAsnGlyThr.....HisHisCysAlaGlySerLeuLeuThr 81
      |||||.....
158 AGATTAAATTCAGCTTCTGGATGATCTTCTCTGGCGGCTCCCTCATTC 207
      |||||.....

```

100

```

268 uatgasatargProglValayrileserleuserlahisargsertry 285
      | ||||| ||||| :||| :||| :||| :||| :||| :||| :|||
763 GGCCAAATCGCTCGCATTTACACCGGGGTGACGTACTACTGACATGGA 812
285 alciulyslleval 289

```

1

[REDACTED]

```

; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-016-366A-20

alignment_scores:
      Quality: 530.50      Length: 288
      Ratio: 2.837      Gaps: 9
      Percent Similarity: 64.931      Percent Identity: 39.236

ment block:
0-040-803-7 x US-09-016-366A-20 ..

Align seg 1/1 to: US-09-016-366A-20 from: 1 to: 1128
20 SerLeuLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAlaArgIlePr 36
   ::::::::::::::::::::
5 AATGTGCTCTCTGTGGCGCTGCTCTG.....GCGAGCCGGCGCTTA 48
   ::::::::::::::::::::
36 oValProProAlaCysGlyLysProGlnGlnLeuAsnArgValAlaGly 53
   ::::::::::::::::::::
49 GCGGCGCCCTGCGCCAGCGCCGCTGCAAGAGTGCGGCGATCGTGGGG 98
   ::::::::::::::::::::
53 LysAlaSerThrAspSerGlyTyrProTyrIleValSerIleGlnLys 69
   ::::::::::::::::::::
99 GTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 148
   ::::::::::::::::::::
70 AsnGly.....ThrHisCysAlaGlySerLeuLeuThrSerAr 83
   ::::::::::::::::::::
149 CAGGCGCATATGAGTGCATCTTGGGGGGCTCCCTCATCCACCCCA 198
   ::::::::::::::::::::
83 GTTPValIleThrAlaAlaHisCys.....PheLysAsp.... 94
   ::::::::::::::::::::
199 GTGGGTGCTGACCGAGCGCACTGCTGGAGACCGAGCGTCAAGATCTGG 248

```

```

95 .....AsnLeuAsnLysProTyrIleuPheSerValLeuLeu 106
   ::::::::::::::::::::
249 CCGCGCTCAGAGGTGCAACTGCGGGAGAGACACTTCTAC..... 286
   ::::::::::::::::::::
107 GAlaIleTyrIleuGlnLysProGlySerArgSerGlnLysValAla 123
   ::::::::::::::::::::
287 .....TACAGAGACCACTGCTGCGCGT 309
   ::::::::::::::::::::
123 IAlaTyrValIleProHisPro...ValTyrSerTyrLysGlnGlyAla 139
   ::::::::::::::::::::
310 CAGCAGATCATGCTGCTACACCATTTACACCGCCAGATCGA... 355
   ::::::::::::::::::::
139 yAlaAspIleAlaLeuValArgLeuGlnArgSerIleGlnPheSerGlu 155
   ::::::::::::::::::::
356 ..GCGGACATCGCCCTGTGAGTGTGAGAGACCGGTGAAGTCTCCAGC 403
   ::::::::::::::::::::
156 ArgValLeuProIleCysLeuProAspAlaSerIleHisLeuProHis 172
   ::::::::::::::::::::
404 CAGTCCACAGGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 453
   ::::::::::::::::::::
172 nThrHisCysTyrIleSerGlyTyrGlySerIleGlnAspGlyValPro 189
   ::::::::::::::::::::
454 GATGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 503
   ::::::::::::::::::::
189 euProHisProGlnThrLeuGlnLysLeuLysValProIleLeuAspSer 205
   ::::::::::::::::::::
504 TCCACCGCGCATTTCTCTGTAAGCAGTGAAGTCCCATATGTAAGAAC 553
   ::::::::::::::::::::
206 GluValCysSerHisLeuTyrTyrArgGlyAlaGlyGlnGlyPro.... 220
   ::::::::::::::::::::
554 CACATTTGTGACGCAAAATACCACTTGGCGCTTACAGCGGAGACGAGCT 603
   ::::::::::::::::::::
221 .....IleThrGluAspMetLeuCysAlaGlyTyrIleuGlnGlyAla 235
   ::::::::::::::::::::
604 CCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 647
   ::::::::::::::::::::
235 rGAspAlaCysLeuGlnLysSerGlyGlyProLeuMetCysGlnValAsp 251
   ::::::::::::::::::::
648 GGGACTCATGCGCAGGCGGCGATCCGAGAGGCGCCCTGCTGCAAGGTAAT 697
   ::::::::::::::::::::
252 GAlaIleTyrLeuAlaGlyIleIleSerTyrGlyGlnGlyCysAlaG 268
   ::::::::::::::::::::
698 GGCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 747
   ::::::::::::::::::::
268 uArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSerTyr 285
   ::::::::::::::::::::
748 GCCCAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 797
   ::::::::::::::::::::
285 alGluLysIleVal 289
   ::::::::::::::::::::
798 TCCACCACTATGTC 811

seq_name: /cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-978-404B-15
seq_documentation_block:
; Sequence 15, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```

[illegible]


```
147 CAGGCCCATCTGATGCATCTTCGCGGGGCTCCCTCATCCACCCCA 196
83 gTTPValIleThrAlaIleHisCys.....PhelyAsp.... 94
:|||||:|||||:|||||:|||||:|||||:
197 GTGGTGCTGACCGCAGCCGACCTGCTGGACCGGACGTCAAGATCTGG 246
95 .....AsnLeuAsnLysProTyrLeuPheSerValLeuLeu 106
:|||||:|||||:|||||:|||||:|||||:
247 CCGCCCTCAGGGGTCAACTGCGGAGCAGACCTCTAC..... 284
107 GLyAlaTyrGlnLeuGlnProGlySerArgSerGlnLysValGlyVa 123
285 .....TACCAGGACCAGCTGCTGCCGT 307
123 lAlaTyrValGluProHisPro...ValTyrSerTyrLysGlnLysAla 139
:|||||:|||||:|||||:|||||:|||||:
308 CAGCAGATCATGCTGACCCACACATTCTACACCGCCCAAGATGGA.... 353
139 ySaLaAspIleAlaLeuValArgLeuGlnArgSerIleGlnPheSerGlu 155
:|||||:|||||:|||||:|||||:|||||:
54 ..GGGACATCGCCCTGCTGAGCTGAGAGCGCGGTGAACGTCTCCAGC 401
156 ArgValLeuProIleCysLeuProAspAlaSerIleHisLeuProAs 172
:|||||:|||||:|||||:|||||:|||||:
402 CAGCTCCACACGGTCACCTGCCCCCTGCTCAGAGACCTTCCCCCGGG 451
172 nThrHisCysTyrIleSerGlyTyrGlySerIleGlnAspGlyValPro 189
:|||||:|||||:|||||:|||||:|||||:
452 GATCCCGTGGTGGTCTGCTGAGGCGCATGTGACAAATGATGAGCGCC 501
189 euProHisProGlnThrLeuGlnLysLeuLysValProIleIleAspSer 205
:|||||:|||||:|||||:|||||:|||||:
502 TCCACCGCCATTCTCTGTGAAGAGGTGAAGTCCCATATGATAAAC 551
206 GluValCysSerHisLeuTyrTyrArgGlyAlaGlyGlnGlyPro.... 220
:|||||:|||||:|||||:|||||:|||||:
552 CACATTGTGTGACGAATAACACCTTGGCGCTACACGGGAGACGACGT 601
221 .....IleThrGlnAspMetLeuCysAlaGlyTyrLeuGlnGlyGln 235
:|||||:|||||:|||||:|||||:|||||:
602 CCGCATGCTCGTGCAGACATGCTGTGCCGG.....AACACCCGGA 645
235 rGaAspAlaCysLeuGlnLysAspSerGlyLysProLeuMetCysGlnValAsp 251
:|||||:|||||:|||||:|||||:|||||:
646 GGGACTCATGCTCAGCGGACCTCCGAGGCGCCCTGCTGCAAGTGAAT 695
252 GLyAlaTyrLeuLeuAlaGlyIleIleSerTyrGlnGlyCysAlaGln 268
:|||||:|||||:|||||:|||||:|||||:
696 GGCACCTGGCTGCAGCGGGGCTGTCAGCTGGGCGAGGGCTGTGCCCA 745
268 uArgAsnArgProGlnValTyrIleSerLeuSerAlaHisArgSerTyr 285
:|||||:|||||:|||||:|||||:|||||:
746 GCCCAACCGGCTGCATACACCGGTGCACCTACTTGACTGGA 795
285 aGlnLysIleVal 289
:|||||
796 TCCACCACTATGTC 809
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:43:12 ; Search time 30.51 Seconds
(without alignments)
1154.061 Million cell updates/sec

Title:	US-10-040-803-7
Perfect score:	1720
Sequence:	1 MVTSGAPPALGGGCTGFTS.....AAGGALRAPSGGSAARS 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

number of hits satisfying chosen parameters: 747574

```

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Database : A:Genesep_032802.*

1:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/hold-genesep/genesep-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1720	100.0	317	20	AAV06482	Human tumour-associ
2	1720	100.0	317	20	AAV13391	Amnio acid sequenc
3	1720	100.0	317	21	AAAB11700	Human serine prote
4	1720	100.0	317	21	AAAY3689	Amnio acid sequenc
5	1720	100.0	317	22	AAAY72850	Human serine prote
6	1720	100.0	317	22	AAAB80259	Human PRO343 prote
7	1720	100.0	325	22	AAAU3215	Human enzyme
8	1720	100.0	325	22	AAU17037	Human novel secret
9	1720	100.0	351	22	AAU16966	Human novel secret
10	1714	99.7	322	22	AAU33751	Novel human enzym
11	1714	99.7	322	22	AAU17038	Human novel secret

12	1706	99.2	319	21	AA11701
13	1636	86.5	306	21	AA11702
14	1484	86.3	271	19	AAW7302
15	1444.5	84.0	327	22	AAV72891
16	1310	76.2	302	21	AA11708
17	1157.5	67.3	308	21	AA11709
18	1074	62.4	234	21	AA11706
19	1047	60.9	207	21	AA11704
20	1017	59.1	201	22	AA75586
21	739	49.0	230	21	AAV7388
22	739	43.0	290	21	AA13938
23	739	43.0	290	22	AAU12282
24	739	43.0	290	22	AAE03821
25	739	43.0	280	22	AAE73945
26	734	42.7	328	22	AA41174
27	730	42.4	327	19	AAW7704
28	680.5	39.6	315	22	AAV7346
29	645	37.5	251	21	AA11705
30	628.5	36.5	284	22	AAU00467
31	585.5	34.0	316	22	AAU00468
32	582.5	33.9	319	21	AA136481
33	582.5	33.9	319	22	AA667541
34	581	33.8	328	21	AA867540
35	581	33.8	328	22	AA867540
36	554	32.2	276	19	AAW64242
37	554	32.2	276	19	AAW63172
38	549	31.9	658	22	AAE06934
39	549	31.9	802	22	AAE06933
40	547	31.8	802	20	AAV41710
41	547	31.8	802	21	AA844266
42	547	31.8	802	21	AA844052
43	538.5	31.3	274	19	AAW64234
44	536.5	31.2	314	19	AAW72977
45	534	31.0	235	22	AAE06932

ALIGNMENTS

RESULT	1	
XX	AA06482	
XX	AA06482 standard; Protein; 317 AA.	
XX	AA06482;	
XX	27-SEP-1999	(first entry)
XX	Human tumour-associated protein PRO343.	
XX	PRO343; UNQ302; cancer; tumour; diagnosis; therapy; human	
XX	Homo sapiens.	
OS		
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..32
FT		/note= "signal peptide"
FT	Protein	33..317
FT		/note= "mature protein"
FT	Modified-site	70
FT		/note= "N-glycosylated"
XX		
PN	WO935170-A2.	
XX		
PD	15-JUL-1999.	
XX		
PF	05-JAN-1999;	99WO-US00106.
XX		
XX	20-NOV-1998;	98US-0109304.
PR	05-JAN-1998;	98US-0070440.
PR	29-APR-1998;	98US-0083500.
PR	22-MAY-1998;	98US-0086414.
PR	10-JUN-1998;	98US-0088742.
PR	10-NOV-1998;	98US-0107783.

PT gastrointestinal ulceration

XX Claim 12; Fig 98; 320pp; English.

CC AAV1344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to
 CC growth or survival of nerve cells including Parkinson's disease,
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.

CC Sequence 317 AA;

Query Match 100.0%; Score 1720; DB 20; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVSGAPPALGGGCGTGTSTLLASTALNARIPVPACGKPOQLNRYVGGEDSTDE 60
 DB 1 MVVSGAPPALGGGCGTGTSTLLASTALNARIPVPACGKPOQLNRYVGGEDSTDE 60
 QY 61 WPIVSIQKNGTHHCAGSLTSRWVITTAACFKDKNLKPIFSVLGAMQOLNGSGSOK 120
 DB 61 WPIVSIQKNGTHHCAGSLTSRWVITTAACFKDKNLKPIFSVLGAMQOLNGSGSOK 120
 QY 121 VGVAVWEHPHYVSWKEGACADIALVRLERSIOFSEVPLPCLPDASIHLPNTHCWISGW 180
 DB 121 VGVAVWEHPHYVSWKEGACADIALVRLERSIOFSEVPLPCLPDASIHLPNTHCWISGW 180
 QY 181 GSIODGVPLPHQDTLOKLVPIIDSEVCSHLWRGAGGPTEDMLCAGYLEGERDACL 240
 DB 181 GSIODGVPLPHQDTLOKLVPIIDSEVCSHLWRGAGGPTEDMLCAGYLEGERDACL 240
 QY 241 DSGGPIKQVNDGAMTLAGITISWEGCAERNRPYVYISLSAHSRWKIVOGVOLGRAG 300
 DB 241 DSGGPIKQVNDGAMTLAGITISWEGCAERNRPYVYISLSAHSRWKIVOGVOLGRAG 300
 QY 301 GGALRAPSGSGAAARS 317
 DB 301 GGALRAPSGSGAAARS 317

RESULT 3
 AAB11700
 ID AAB11700 standard; Protein; 317 AA.

XX AAB11700;
 XX 23-OCT-2000 (first entry)

XX Human serine protease BSSP4 (hBSSP4) SEQ ID NO:2.

XX BSSP4; serine protease; human; hBSSP4; mouse; mBSSP4; brain;
 KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;
 KW oedema; dropsy; cancer; inflammation; prostate; testis; bone.

OS Homo sapiens.

XX WO200031277-A1.

XX 02-JUN-2000.

XX 19-NOV-1999; 99WO-JP06472.

XX 20-NOV-1998; 98JP-0347813.

XX (FUSO) FUSO PHARM IND LTD.

XX Uemura H, Okui A, Komimami K, Yamaguchi N, Mitsui S;

XX WPI; 2000-400084/34.

PT Serine protease BSSP4 and antibodies recognizing BSSP4 for assay and
 PT diagnosis of diseases in which BSSP4 expression is altered

PS Claim 1; Page 66-67; 11pp; Japanese.

CC The invention relates to novel serine proteases designated BSSP4
 CC (AAB11700-B11709), and to nucleic acids encoding them (AAA61695-A61704,
 CC AAA61799). The invention also relates to vectors and transformants
 CC comprising BSSP4 nucleic acids; transgenic animals in which the
 CC expression level of BSSP4 can be varied; and an mBSSP4 knockout mouse.
 CC The invention additionally encompasses anti-BSSP4 antibodies and methods
 CC of production of such antibodies, methods of BSSP4 detection using the
 CC antibodies, and the use of BSSP4 proteins or fragments as diagnostic
 CC markers for certain medical conditions. Nucleotides encoding BSSP4 were
 CC initially isolated in a human brain cDNA library using degenerate PCR
 CC primers (AAA61714-A61715) based on conserved regions of serine
 CC proteases. The BSSP4 serine proteases and nucleotides encoding them are
 CC useful in detecting homologues, mutants and polymorphic variants in
 CC biological samples (e.g., blood, urine, brain, prostate gland and testis)
 CC as diagnostic markers for diseases associated with altered BSSP4
 CC expression levels. Such diseases include Alzheimer's disease, oedema
 CC (dropsy), cancer or inflammation of brain, prostate, testis or bone.
 CC Sequences AAA61695-A61703 and AAA61799 represent cDNAs encoding human
 CC BSSP4 variants (hBSSP4), and sequence AAA61704 represents cDNA encoding
 CC murine BSSP4 (mBSSP4). Sequences AAB11700-B11708 represent human BSSP4
 CC variants (hBSSP4), and sequence AAB11709 represents murine BSSP4
 CC (mBSSP4).

CC Sequence 317 AA;

Query Match 100.0%; Score 1720; DB 21; Length 317;
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVSGAPPALGGGCGTGTSTLLASTALNARIPVPACGKPOQLNRYVGGEDSTDE 60
 DB 1 MVVSGAPPALGGGCGTGTSTLLASTALNARIPVPACGKPOQLNRYVGGEDSTDE 60
 QY 61 WPIVSIQKNGTHHCAGSLTSRWVITTAACFKDKNLKPIFSVLGAMQOLNGSGSOK 120
 DB 61 WPIVSIQKNGTHHCAGSLTSRWVITTAACFKDKNLKPIFSVLGAMQOLNGSGSOK 120
 QY 121 VGVAVWEHPHYVSWKEGACADIALVRLERSIOFSEVPLPCLPDASIHLPNTHCWISGW 180
 DB 121 VGVAVWEHPHYVSWKEGACADIALVRLERSIOFSEVPLPCLPDASIHLPNTHCWISGW 180
 QY 181 GSIODGVPLPHQDTLOKLVPIIDSEVCSHLWRGAGGPTEDMLCAGYLEGERDACL 240
 DB 181 GSIODGVPLPHQDTLOKLVPIIDSEVCSHLWRGAGGPTEDMLCAGYLEGERDACL 240
 QY 241 DSGGPIKQVNDGAMTLAGITISWEGCAERNRPYVYISLSAHSRWKIVOGVOLGRAG 300
 DB 241 DSGGPIKQVNDGAMTLAGITISWEGCAERNRPYVYISLSAHSRWKIVOGVOLGRAG 300
 QY 301 GGALRAPSGSGAAARS 317
 DB 301 GGALRAPSGSGAAARS 317

XX	New anti-polypeptide antibody useful in the treatment and diagnosis of neoplastic cell growth and proliferation -
PT	
XX	Claim 61; Fig 12; 220bp; English.
PS	
XX	The present sequence represents a novel human polypeptide. The specification describes novel polypeptides designated PRO201, PRO292, CC PRO327, PRO1265, PRO344, PRO343, PRO347, PRO357, PRO715, PRO1017, CC PRO1112, PRO5059, PRO853 and PRO882. These genes are amplified in the genome of tumour cells. The polypeptides are believed to contribute CC to tumorigenesis. The polypeptides are useful target for the CC identification of certain cancers, and may act as predictors of the CC prognosis of tumor treatment. Antibodies against these polypeptides CC are useful in the treatment and diagnosis of neoplastic cell growth CC and proliferation in mammals.
CC	
XX	
SQ	Sequence 317 AA;
	Query Match 100.0%; Score 1720; DB 21; Length 317; Best Local Similarity 100.0%; Pred. No. 3.2e-147; Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 MVVSGAPPALGGCGCTGFTSLLLASTAIIINAAIPVPACGKPOOLNRYVGEDSTDSSE 60 Db 1 mvvsgappalggcgctgftslallastalnaaripvpacgkpgqlnrvggedstde 60
OY	61 WPIWVISIÖKNKGTHHCAGSILTSRWYTIAHCFKDMLNKPYLFVSULLGAMOLGNPGSRÖK 120 Db 61 wpwivsiöknkgthhcagsiltsrwytiaahcfkdmlnkpylfsullgwqjgnpsrsqk 120
OY	121 VGVAVVEHPHYVMKEGCADIALVRLERSIOFSERNVIPICLPDASIHLPPTHGWISGM 180 Db 121 vgvavvehpptyvmkegcadialvrleersiofservipiclpdasihlpptchwisgm 180
OY	181 GSIGDPVLPHPQTLOKLKVPIIDSEWCSHLYMRGAGOGPTEDMLCAGYLEGERDACLG 240 Db 181 gsigdgvplphpqtlöklkvpiidsewcshlymrpagogptedmlcagylegerdaclg 240
OY	241 DSGGIMQOVNGAWILLAGTIISMGECCARRNPGYTISIASHRSWEKYIVQVOLRGRAOG 300 Db 241 dsggimqvngawillaglitsmgecarrnprgytisishrswekyivqvölgragrög 300
OY	301 GGALRAPSQSGGAARS 317 Db 301 ggaltapsgsggaars 317
RESULT	5
AAAY72890	AAAY72890 standard; Protein; 317 AA.
XX	AAV72890;
XX	
DF	31-MAY-2001 (first entry)
XX	
DE	Human serine protease, protease C-E.
KW	Human: serine protease; protease C-B; therapy; desquamation; skin care; KW laundry detergent; shampoo; cleaning agent; hair care; skin flaking; KW neurodegenerative disorder; dermatological; immunogenic; proteolytic. XX chromosome 16p13.3.
Homo sapiens.	
M0200116288-A2.	
08-MAR-2001.	
14-AUG-2000; 2000MO-US22117.	
99MS-0786539	

Human; PBO; dermatological; antipsoriatic; cytostatic; antinflammatory;
antiparkinsonian neurotropic; neuroprotective; valneray; cardiac;
antiangiogenic; vasotropic; antiasomatic; antirheumatic; cancer;
antiarthritic; antifertility; antidiabetic; antiviral; diabetes;

121 VGVAVVEPHPIVISMKEGACADIAL

|||||
Db 121 ygvawephpyswkegcadialvrlersigfseervlplclpaslhlppnthcwisgw 180
QY 181 GSIDGVLPHRPQTLQKAVIITISEVCSHLYMRAGGPTTEMLCAGYEGEERDCLG 240
Db 181 gsidgvlphrpqtlqkavilisevcsshlymragsptteemlcagylegerdclg 240
QY 241 DSGGFLMCQVGGAMLIIAGIISWEGSCAERNRPVYISLSAHSRWEKIVQGLRGAQC 300
Db 241 dsggflmcqvvgawllagliswgegcacernrpyvylsahsrwekivgqlrgrag 300
QY 301 GGALRAPSQSGGAARS 317
Db 301 ggalrapsqsgsaaars 317
RESULT 7
AAU23215
AAU23215 standard; Protein; 325 AA.
AAU23215;
17-DEC-2001 (first entry)
XX
XX Novel human enzyme polypeptide #301.
DE
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
XX Homo sapiens.
OS
XX MO200155301-A2.
PN
XX 02-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001MO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 27-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0218293.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226661.

PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229513.
PR 05-SEP-2000; 2000US-0229519.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239355.
PR 13-OCT-2000; 2000US-0239357.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 03-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-465566/50.
 DR N-PSDB: AAS41085.
 XX

Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases.

Claim 11; SEQ ID NO 1211; 1180bp; English.

The present invention relates to the isolation of novel human enzyme polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy. The AAT2915-AAT23814 represent the novel human enzyme polypeptides of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 325 AA:

Query Match 100.0%; Score 1720; DB 22; Length 325;
 Best Local Similarity 100.0%; Pred. No. 3.3e-147;
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVSGAPPALGGCLGTFSTLLASTALNAAIPVPACGKPOOLNRVVGEDSTDSE 60
 |||||||
 Db 9 mvysgappalggcigtftsl|||lastal|naar|pvpacgkppq|nrvvgedstdse 68
 QY 61 WPIVSIQKNGTHCAGSLTSSRWVITAAHCFKDLNKPYLESVLLGAMQIGNGSSQK 120
 |||||||
 Db 69 wpwivsiqkngthcagslltssrwvitaahcfkdl|nkpyle|svll|gawq|gnpgrsqk 128
 QY 121 VGVAWEHPHYSMKEGACADIALVRLERSIOFSRVPICLPASTHLPPNTHQWISGW 180
 |||||||
 Db 129 vgvawehphysmwkegcadialvrlersiofserv|piclpas|hl|pnt|hqw|isgw 188
 QY 181 GSIDGVPPLPHTLOKLPVPIIDSEVCSHLVNRAGGQPTTDMTCAGYLEGERDACLG 240
 |||||||
 Db 189 gsidgvp|lph|t|lq|k|vp|i|d|s|e|v|c|s|h|l|v|n|r|a|g|g|p|t|t|d|m|t|c|a|g|y|l|e|g|e|r|d|a|c|l|g 248
 QY 241 DSGGPLMCOVDGAMTLNGLITSMGSCAERNRPYIISAHRSVEXEIVGVQLRGAQG 300
 |||||||
 Db 249 dsggplmcvqvgaw|l|n|g|l|t|s|m|g|s|c|a|e|r|n|p|y|i|s|a|h|r|s|v|e|x|e|i|v|g|v|q|l|r|g|a|q 308
 QY 301 GGALRAPSGSGAARS 317
 |||||||
 Db 309 ggalrapsgsgsaars 325

RESULT 8

AAU17037 standard; Protein: 325 AA.

AAU17037:

07-NOV-2001 (first entry)

Human novel secreted protein, SEQ ID 278.

Human: immunosuppressive; antiarthritic; antirheumatic;
 cytostatic; cardiant; vasotropic; cerebroprotective; mototropic;
 neuroprotective; antibacterial; vitricide; fungicide; optalmological;
 vunerary; secreted protein; rheumatoid arthritis;
 hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
 cerebrovascular disorder; cerebral ischaemia; angiogenesis;
 nervous system disorder; Alzheimer's disease; infection; ocular disorder;
 corneal infection; wound healing; epithelial cell proliferation;
 skin ageing; food additive; preservative; antiproliferative.

Homo sapiens.

WO200155441-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US01320.

PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.

disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capability, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present

Query Match 100.0%; Score 1720; DB 22; Length 325;
Best Local Similarity 100.0%; Pred. No. 3.3e-147;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MVSAGAPALGGGCLGTFSTLLASTATLAAARIPVPACGKPOQINRVYGGEDSTDE 60
|||||
9 mvysagapalgggclgftstllastatlaaariipvpacgkpgqlnrvyggedstde 68
61 WPIVTSIQKNGTHHCAGSLTSMVITANHCDFDNLNKPYLFESVLLGAWQLGMPGSRQK 120
|||||
69 wpivtsiqkngthhcagsltsmvitahcfdnlnkpylfesvllgawqlgmpgsrsk 128
121 VGVAVWEPHPYVSWKEGACADIALVRLERSIOFSESVLPICLPDASIHLPNTHCWISGW 180
|||||
129 vgvawephpyswkcgacadiavrlersiofsevsrlypiclpdasihlpnthcwisgw 188
181 GSIDGVPLPHTQTLKRLKVPIDSEVCSHLVWRGACGPTEDMTCAGLSEGRDACLG 240
|||||
189 gsldgvplphtqlkrlkvpidsevcshlvwrgacgptedmtcagylsegerdclg 248
241 DSGPIMCQVNGAFLAGITISWEGCAERNRPYVYISLSHRSVWEKIVQGVOLRGRAQG 300
|||||
249 dsgpimcqvgawflagitiswegcaernrpyvylsishrsvekivqgvqlrgraq 308
301 GGALRAPSOGSGAARS 317
|||||
309 ggalrapsogsgaars 325

RESULT 9
AAU16966
ID AUI16966 standard; Protein; 351 AA.
XX
AC AUI16966;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, SEQ ID 207.
Human; immunosuppressive; antiarthritic; antirheumatic;
cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
neuroprotective; antibacterial; virucide; fungicide; optalmatological;
vulnerable; secreted protein; rheumatoid arthritis;
hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
cerebrovascular disorder; cerebral ischemia; angiogenesis;
nervous system disorder; Alzheimer's disease; infection; ocular disorder;
corneal infection; wound healing; epithelial cell proliferation;
skin ageing; food additive; preservative; antiproliferative.

OS Homo sapiens.
XX
XX
PN WO20015441-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01320.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 11-JUL-2000; 2000US-0218299.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227909.
PR 30-AUG-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.

	Query Match	100.0%	Score 1720;	DB 22;	Length 351;
	Best Local Similarity	100.0%	Pred. No. 3.6e-147;		
	Matches 317;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 MVSAGAPALGGCGIGFTSLLASTLAFTLAARIPYPACGRPOOLNRYVGGEDSTDSE 60				
Db	35 mvvgagpapaigggclgfttsllllastcalnaatirpypacgkbpqinrvyvgedsdse 94				
QY	61 WPWIVSIQKNGTTHHCAGSLTSMRWITPAHCFKDNLKPYLEFVLLGAMOLGNGSRSQK 120				
Db	95 wpwivsiqkngthhcagsltswrwitlaahcfkdnlkpylfsvllgawqgmprsqk 154				
QY	121 VGVAMVEHPHYSMKKEGACADIALVLESTIOSEFRLPCLPASTHLPENTHCWISGW 180				
Db	155 vgvamvehpnysvkqgcadialvlersqtserylplcldpaslnlpntcwisgw 214				
QY	181 GSIDGVPLEPHQYLOKILKYPIDISEVCSHLMYRGAOGPITDEMLCAGYLEGERDACLG 240				
Db	215 gsidgvpplphqclqklkypidisevcslnhyvrgagqptledmclcagylegerdclg 274				
QY	241 DSGGPIMCOUNGAWLILAGIISMGEGCAERNRPYVITSLASRWSVETIVQCYOLRGARAG 300				
Db	275 dsggplmcqvqgawllagilswgegcaernpyvilsanrswvexkivqvgqltraga 334				
QY	301 GGALRAPSGSGGAARS 317				
Db	335 ggalrapsgsggaars 351				
RESULT 10					
AAU23751					
AAU23751	AAU23751 standard; Protein; 322 AA.				
XX	AAU23751:				
DT	18-DEC-2001 (first entry)				
DE	Novel human enzyme polypeptide #837.				
XX					
KM	Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;				
KW	ligase; hyperproliferative disorder; immunodeficiency disorder;				
KM	autoimmune disorder; neurological disorder; metabolic disorder;				
KW	inflammatory disorder; cardiovascular disorder; reproductive disorder;				
KM	blood-related disorder; infectious disorder; cytostatic; anti arthritic;				
KM	nephrotropic; anticoagulant.				
OS	Homo sapiens.				
PN	WO200155301-A2.				
XX	02-AUG-2001.				
XX					

PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0236830.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237039.
PR	13-OCT-2000;	2000US-0237340.
PR	13-OCT-2000;	2000US-0239363.
PR	13-OCT-2000;	2000US-0239367.
PR	20-OCT-2000;	2000US-0239560.
PR	20-OCT-2000;	2000US-0241121.
PR	20-OCT-2000;	2000US-0241121.
PR	20-OCT-2000;	2000US-0241186.
PR	20-OCT-2000;	2000US-0241187.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	17-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249219.
PR	17-NOV-2000;	2000US-0249220.
PR	17-NOV-2000;	2000US-0249221.
PR	17-NOV-2000;	2000US-0249222.
PR	17-NOV-2000;	2000US-0249223.
PR	17-NOV-2000;	2000US-0249224.
PR	17-NOV-2000;	2000US-0249225.
PR	17-NOV-2000;	2000US-0249226.
PR	17-NOV-2000;	2000US-0249227.
PR	17-NOV-2000;	2000US-0249229.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250351.
PR	05-DEC-2000;	2000US-0250351.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251198.
PR	05-DEC-2000;	2000US-0251719.
PR	06-DEC-2000;	2000US-0251747.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	Rosen CA, Barash SC, Ruben SM,	
XX	WPI; 2001-465566/50.	
XX	N-PSDS; AAS41621.	

XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -

XX
PS Claim 11: SEQ ID No 1747; 1180bp; English.

XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AA040785-AA041684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders, including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS), autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AA022915-AA023814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 322 AA;

Query Match 99.7%; Score 1714; DB 22; Length 322;

Best Local Similarity 99.7%; Pred. No. 1.1e-146; Indels 0; Gaps 0;

Matches 316; Conservative 0; Mismatches 1;

QY 1 MVSAGAPALGCGCTGFTSILLASPAIILAAIRIPVPAAGKPOOLNRYVGGEEDSTDE 60
DB 6 MVSAGAPALGCGCTGFTSILLASPAIILAAIRIPVPAAGKPOOLNRYVGGEEDSTDE 65
QY 61 WPIVSIKNGTTHCAGSILTSRWVITAHCFKDNLRKPYLFVILGAMOLGNPSRSQK 120
DB 66 WPIVSIKNGTTHCAGSILTSRWVITAHCFKDNLRKPYLFVILGAMOLGNPSRSQK 125
QY 121 VGVAVNEPYPYVSKEGACADIALVRLERSTQSERVLPICLPDASTIHLPPNTHCWNISGW 180
DB 126 VGVAVNEPYPYVSKEGACADIALVRLERSTQSERVLPICLPDASTIHLPPNTHCWNISGW 185
QY 181 GSTIDGVPLPHPOTLQKLVPIIDSEVCSHLVWRAGAGPITTEDMLCAGYLEGERDACLG 240
DB 186 GSTIDGVPLPHPOTLQKLVPIIDSEVCSHLVWRAGAGPITTEDMLCAGYLEGERDACLG 245
QY 241 DSGGPLMCCVDGAWLLACIISWEGCARRNPYIISASHSWVEKTIYGVOLRGRAOG 300
DB 246 DSGGPLMCCVDGAWLLACIISWEGCARRNPYIISASHSWVEKTIYGVOLRGRAOG 305
QY 301 GGATRAPSGSGAARS 317
DB 306 GGATRAPSGSGAARS 322

Search completed: August 13, 2002, 08:47:16
Job time: 244 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: August 13, 2002, 08:45:47 ; Search time 17.42 Seconds
(without alignments)
1748.583 Million cell updates/sec

Title: US-10-040-803-7

Perfect score: 1720

Sequence: 1 MYVSGAPPALGGGCLGRTFS.....AOGGALRAPSGGGAARS 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	607.5	35.3	343	1 A57014	proctasin (EC 3.4.21.1)
2	554	32.2	276	2 A38654	mast cell proteina
3	546	31.7	275	2 A32410	trypsin (EC 3.4.21.1)
4	538.5	31.3	274	2 JC4171	trypsin (EC 3.4.21.1)
5	536	31.2	237	2 S68702	trypsin (EC 3.4.21.1)
6	532	30.9	274	2 A45754	trypsin (EC 3.4.21.1)
7	530.5	30.8	275	2 B35863	trypsin (EC 3.4.21.1)
8	529.5	30.8	275	2 A35863	trypsin (EC 3.4.21.1)
9	522.5	30.4	275	2 C35863	trypsin (EC 3.4.21.1)
10	522	30.3	270	2 S56160	mast cell trypsin
11	518	30.1	273	2 A47246	trypsin (EC 3.4.21.1)
12	481	29.6	638	1 KOHUP	plasma kallikrein
13	475.5	29.6	625	1 KFHU1	coagulation factor
14	473	27.3	638	1 KOMSP	plasma kallikrein
15	469.5	27.3	638	1 KORTPL	plasma kallikrein
16	460.5	26.8	812	1 PLBO	plasma (EC 3.4.21.1)
17	457.5	26.6	1034	1 A53663	enteropeptidase (E
18	452	26.3	417	1 S00845	hepsin (EC 3.4.21.1)
19	450	26.2	416	1 S37777	hepsin (EC 3.4.21.1)
20	450	26.2	810	2 B30848	plasma (EC 3.4.21.1)
21	446	25.9	455	2 A61545	plasma (EC 3.4.21.1)
22	446	25.9	790	1 PLPG	plasma (EC 3.4.21.1)
23	446	25.9	810	1 PLHU	polypeptide - Afri
24	445.5	25.9	1524	2 T30337	serine proteinase
25	445	25.9	786	1 A47547	enteropeptidase (E
26	444	25.8	1019	1 A56318	enteropeptidase (E
27	443.5	25.8	1035	1 A45090	apolipoprotein(a)
28	442	25.7	1420	2 A32869	plasma (EC 3.4.21.1)
29	441	25.6	460	2 B61545	plasma (EC 3.4.21.1)

30	439.5	25.6	812	1 PLMS	plasma (EC 3.4.21.1)
31	437.5	25.4	367	2 JF0104	testicular serine
32	437.5	25.4	437	2 S18407	acrosin (EC 3.4.21.1)
33	437.5	25.4	4548	1 S00657	apoptein(a) (EC 3.4.21.1)
34	437	25.4	263	1 KYRFB	chymotrypsin (EC 3.4.21.1)
35	437	25.4	421	1 S11674	acrosin (EC 3.4.21.1)
36	434	25.2	265	2 T15451	hypothetical prote
37	433.5	25.2	271	2 A25528	pancreatic elastas
38	432.5	25.1	436	2 JX0172	acrosin (EC 3.4.21.1)
39	430.5	25.0	366	2 JE0105	testicular serine
40	429.5	25.0	431	2 S47538	acrosin (EC 3.4.21.1)
41	428	24.9	263	2 A21195	chymotrypsin (EC 3.4.21.1)
42	425	24.7	810	2 I46260	plasma (EC 3.4.21.1)
43	424.5	24.7	270	2 B29934	pancreatic elastas
44	422	24.5	263	2 A31299	chymotrypsin (EC 3.4.21.1)
45	422	24.5	415	1 A34170	acrosin (EC 3.4.21.1)

ALIGNMENTS

RESULT 1

A57014
proctasin (EC 3.4.21.1) precursor - human
C:Species: Homo sapiens (man)
C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 18-Jun-1999
C:Accession: A57014; A54866
R:Yu, J.X.; Chao, L.; Chao, J.
J. Biol. Chem. 270, 13483-13489, 1995
A:Title: Molecular cloning, tissue-specific expression, and cellular localization of
A:Reference number: A57014; M01D:95286644
A:Accession: A57014
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-343 <RES>
A:Cross-references: GB:U41351; NID:9862304; PIDN:AA041759.1; PID:9862305
A:Experimental source: prostate
A:Note: parts of this sequence were determined by protein sequencing
R:Yu, J.X.; Chao, L.; Chao, J.
J. Biol. Chem. 269, 18843-18848, 1994
A:Title: Proctasin is a novel human serine proteinase from seminal fluid. Purificatio
A:Reference number: A54866; M01D:94308140
A:Accession: A54866
A:Molecule type: protein
A:Residues: 45-64 <YUA>
C:Genetics:
A:Gene: GDB:PRSS8
A:Cross-references: GDB:676446; OMIM:600823
A:Map position: 16p11.2-16p11.2
C:Superfamily: proctasin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-44/Domain: proctasin #status predicted <CHL>
F:45-343/Domain: proctasin light chain #status predicted <CHL>
F:45-343/Domain: proctasin heavy chain #status predicted <CHH>
F:45-281/Domain: trypsin homology <TRY>
F:33-341/Domain: transmembrane #status predicted <TM1>
F:37-154, 70-86, 168-244, 201-223, 234-262/disulfide bonds: #status predicted
F:85,134,238/Active site: His, Asp, Ser #status predicted
F:159/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 35.3% Score 607.5; DB 1; Length 343;

Best local similarity 44.2% Pred. No. 6.3e-44;
Matches 129; Conservative 42; Mismatches 108; Indels 13; Gaps 7;

QY 10 LGGGCLGTFSTL---LIASTAILNARIPIVPPACGRPOQNRVVGSGEDSDSEMPVVS 66
DB 7 LGGPGLGAVAILLYLGLRSGTGAEGAEAP---CGVAPQA-RTTGGSSAVAGQMPQVS 61
QY 67 IQKNTHHCAGSLILSRVITTAHOFKDNLNKPYLFVSLGAMOLNGSGSRQKGVAVV 126
DB 62 ITYESVHVCGSLVSEQWVLSAHCFFSEHKE-AIEYKGLAHQLDSTSEDAKYSTLKD 120

QY 127 EPHVYSKREACADIALAVTLERSTOSENRLPTCLPDASIHLPNNHCHISGWSITDGB 186
 Db 121 IPHBSY-LQESSGCDIALLOSRPTTSRYKTRICPEAANASFPNGHCHVTWGHVAPS 179
 QY 187 VPLPHPTOLKRLKVPYIDSEVCSHLWYRGA-GGSP--ITEDMLCAGLEGEBRDACIGDSG 243
 Db 180 VSLTPKRLQGLEVPILSRFCNCLVIYDAKPEEPHFVQEDMVCAGYVEBGKACGGDSG 239
 QY 244 GPLMCOVDGAULLAGIISWBGCAERNRPVYIISLASHRSEVKEIYOGVOLF 295
 Db 240 GPLSCPVGMLYTLGIYVSWGDACAGARRRPVYITLASSYASWIDSKYTELOPR 291

RESULT 2

mast cell proteinase 6 (EC 3.4.21.-) precursor - mouse
A38654

C.Species: Mus musculus (house mouse)
C.Date: 21-Feb-1992 #sequence revision 17-Feb-1994 #text_change 22-Jun-1999
Accession: A38654; B38654; D35646; I59478
Hendols, D.S.; Gurley, D.S.; Austen, K.F.; Serafin, W.E.
Mol. Chem. 266, 3847-3853, 1991
Title: Cloning of the cDNA and gene of mouse mast cell protease-6. Transcription by p

A.Reference number: A38654; MUID:91139682
A.Accession: A38654
A.Molecule type: DNA
A.Residues: 1-276 <REY>
A.Cross-references: GB:M57625; NID:g200506; PIDN:AAA39987.1; PID:g200507
A.Note: The authors translated the codon CGC for residue 24 as Ala, GAG for residue 37 as Gly, GAG for residue 148 as Gly, GAG for residue 168 as Gly, and GAA for 185 as Gly
A.Accession: B38654
A.Molecule type: mRNA
A.Residues: 1-276 <RE2>
A.Cross-references: GB:M57626; NID:g200508; PIDN:AAA39988.1; PID:g200509
R.Reynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Serafin, W.E.
Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990
A.Title: Different mouse mast cell populations express various combinations of at least
A.Reference number: A35646; MUID:90222202
A.Accession: D35646
A.Molecule type: Protein
A.Residues: 32-54 <RES>
A.Molecule type: mRNA
A.Residues: 1-276 <RES>
A.Cross-references: GB:L31853; NID:g473480; PIDN:AAA39725.1; PID:g473481

Genetics:

A.Gene: MMCP-6
A.Introns: 24/1; 79/2; 168/1; 222/3
A.Superfamily: trypsin; trypsin homology
C.Keywords: hydrolase; serine proteinase; zymogen
F.1-21/Domain: signal sequence #status predicted <SIG>
F.22-31/Domain: activation peptide #status predicted <ACT>
F.32-276/Product: mast cell proteinase 6 #status experimental <CAT>
F.33-268/Domains: trypsin homology <TRI>
F.75,122,225/Active site: His, Asp, Ser #status predicted

[illegible]

```

RESULT      3
A32410
trypsinase (EC 3.4.21.59) precursor - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 22-Jun-1999
C:Accession: A32410
R:Vandersilice, P.; Craik, C.S.; Nadel, J.A.; Caughey, G.H.
Biochemistry 28, 4148-4155, 1989
A:Title: Molecular cloning of dog mast cell trypsinase and a related protease: structure
A:Reference number: A32410; MUID:89352460
A:Accession: A32410
A:Molecule type: mRNA
A:Residues: 1-275 <VNA>
A:Cross-references: GB:M24664; NID:G163982; PIDN:AAA30854.1; PID:G163983; GB:J02862
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Product: trypsin #status predicted <MOT>
E:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted

```

	Query Match	32.8;	Score 554;	DB 2;	Length 276;	
	Best Local Similarity	41.68;	Pred.No.1.7e-39;			
	Matches 116;	Conservative 48;	Mismatches 93;	Indels 22;	Gaps 9;	
OY	21 LLLASTAIIINARIPVP-PACGAPQOLNRYVGEGDSTSEWMIYSIQ---KNGHHCA 76 : : : : Db 6 LLLMALSLASLYVSAPRPA-----NORGVIGSGHESASESKMWOVSLFKINIWIHECG 61 OY 77 GSIILTSRMVITTAACHFCNDLNRPFLFSVLIGAMQL--GNPGRSRQKVAGVAEPPHPVYSM 134 Db 62 GSLLHPQMVLTAACHVCSPHIKSQPLRFYGLRDGYLYGG----QLSLNRIVVPHPHYTT 116					
	RESULT 4					
	JC4171	tryptase (EC 3.4.21.59) precursor - rat				
	N:Alternate names:	mast cell tryptase				
	C:Species:	Rattus norvegicus (Norway rat)				
	C>Date:	Aug-1995 #sequence_revision 27-Oct-1995 #text_change 20-Jun-2000				
	C:Accession:	JC4171				
	R:Ido, H.; Toho, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Matuyama, H.; Osada, Y.	Biochem 118, 210-215, 1995				
	A:Title:	cDNA sequencing and expression of rat mast cell tryptase.				
	A:Reference number:	JC4171; PMID:96015171				

RESULT 7

B35863

tryptase (EC 3.4.21.59) II precursor - human

N:Alternate names: tryptase beta

C:Species: Homo sapiens (man)

C>Date: 03-Feb-1994 #sequence, revision 03-Feb-1994 #text_change 21-Jul-2000

C/Accession: B35863; A37193; I59473

R:VanderSluis, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.

Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990

A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr

A:Reference number: A35863; MUID:90251647

A:Accession: B35863

A:Molecule type: mRNA

A:Residues: 1-275 <VAN>

A:Cross-references: GB:M33492; NID:g339982; PIDN:AAA36779.1; PID:g339983

A:Note: residues 2-275 are derived from mRNA; residue one was inferred from the genomic

R:Miller, J.S.; Moxley, G.; Schwartz, L.B.

C:Clon. Invest. 86, 864-870, 1990

A:Title: Cloning and characterization of a second complementary DNA for human tryptase.

A:Reference number: A37193; MUID:90369005

A:Accession: A37193

A:Molecule type: mRNA

A:Residues: 1-275 <MIT>

A:Cross-references: GB:M37488; NID:g179583; PIDN:AAA51843.1; PID:g179584

R:Blom, T.; Hellman, L.

Scand. J. Immunol. 37, 203-208, 1993

A:Title: Characterization of a tryptase mRNA expressed in the human basophil cell line K

A:Reference number: I59473; MUID:9166209

A:Accession: I59473

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-275 <RES>

A:Cross-references: GB:S55551; NID:g265666; PIDN:AAD13876.1; PID:g4261576

A:Experimental source: basophil cell line KU812

A:Genetics:

A:Gene: GDB:TPS1

A:Cross-references: GDB:125890; OMIM:191080

A:Map position: 16pter-16qter

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-30/Domain: activation peptide #status predicted <ACT>

F:31-275/Product: trypsin I #status predicted <MAT>

F:31-267/Domain: trypsin homology <TRY>

F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 30.8%; Score 530.5; DB 2; Length 275;

Best Local Similarity 39.2%; Pred. No. 1.6e-37;

Matches 113; Conservative 53; Mismatches 85; Indels 37; Gaps 9;

DB 20 SLTLASTALINMARIPVPACGKPOQLRVGSDSTSEMPVIVSIQKNG---THHCA 76

DB 3 NLLILALPVL--ASRAVAAPAGALORVIGVGOEAPRSKRPVQVSLKVGKPTVMHFCG 60

DB 77 GSLLTSRWVITAAHC---FKD-----NLNKPYLESVLLGAMOLGNGSRQKGVAVY 126

DB 61 GSLHPQWVLTAAHCVGRDVKDLAALRVQLREQHLV-----YDQQLLPVSRI 107

DB 127 EPRP-VYSWKEGACADIALVRLRSIOFSERVLPICLPDASIHLPNTNHCISGKSID 185

DB 108 IVHPOFTYTAQIG--ADIALLELEPVSNNHVTTLPPASETPRPQPCWVTGMDVDN 165

DB 186 GVPLRPQTLQKLVPIIDSEVCSHLVWKGAGGPR---ITEDMTCAGYLEGEBDACLGD 241

DB 166 DERLPFRPLQKQKVVIMENHICDAKYHGLGAYTGDDVRIYRDMILCAG--NTRRDSQGD 223

DB 242 SGGPLKQVDAWLLAGTISWEGCAERNRPQVYISLSAHRSWVEKTV 289

DB 224 SGGPLVCKVNGTWLDAGVSWGEGCAQPNRPQYITRVYLYDMTHHYV 271

RESULT 8

A35863

tryptase (EC 3.4.21.59) I precursor - human

C:Species: Homo sapiens (man)

C>Date: 12-Oct-1990 #sequence, revision 12-Oct-1990 #text_change 21-Jul-2000

C/Accession: A35863; D35863; A60939; A39326

R:VanderSluis, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.

Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990

A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine

A:Reference number: A35863; MUID:90251647

A:Accession: A35863

A:Molecule type: DNA

A:Residues: 1-275 <VAN>

A:Cross-references: GB:M33494; NID:g3927804; PIDN:AAC83172.1; PID:g339977

A:Accession: D35863

A:Molecule type: mRNA

A:Residues: 1-275 <VA2>

A:Cross-references: GB:M33491

R:Butterfield, J.H.; Weller, D.A.; Hunt, L.W.; Wynn, S.R.; Roche, P.C.

J. Leukoc. Biol. 47, 409-419, 1990

A:Title: Purification of tryptase from a human mast cell line.

A:Reference number: A60939; MUID:90244210

A:Accession: A60939

A:Molecule type: protein

A:Residues: 31-38, 'P', '40-41', 'X', '43', 'T', '45-48', 'X', '50' <BUT>

A:Experimental source: mast cell

A:Note: 44-Gly was also found

R:Cromlish, J.A.; Seidah, N.G.; Marcinkiewicz, M.; Hamelin, J.; Johnson, D.A.; Chretil

J. Biol. Chem. 262, 1363-1373, 1987

A:Title: Human pituitary tryptase: molecular forms, NH-2-terminal sequence, immunocy

A:Reference number: A39326; MUID:87109258

A:Accession: A39326

A:Molecule type: protein

A:Residues: 31-38 <CRO>

A:Experimental source: pituitary

C:Genetics:

A:Insertions: 21/1; 78/2; 177/1; 221/3

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-30/Domain: activation peptide #status predicted <ACT>

F:31-275/Product: trypsin I #status experimental <MAT>

F:31-267/Domain: trypsin homology <TRY>

F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 30.8%; Score 529.5; DB 2; Length 275;

Best Local Similarity 39.2%; Pred. No. 2e-37;

Matches 113; Conservative 52; Mismatches 86; Indels 37; Gaps 9;

DB 20 SLTLASTALINMARIPVPACGKPOQLRVGSDSTSEMPVIVSIQKNG---THHCA 76

DB 3 NLLILALPVL--ASRAVAAPAGALORVIGVGOEAPRSKRPVQVSLKVGKPTVMHFCG 60

DB 77 GSLLTSRWVITAAHC---FKD-----NLNKPYLESVLLGAMOLGNGSRQKGVAVY 126

DB 61 GSLHPQWVLTAAHCVGRDVKDLAALRVQLREQHLV-----YDQQLLPVSRI 107

DB 127 EPRP-VYSWKEGACADIALVRLRSIOFSERVLPICLPDASIHLPNTNHCISGKSID 185

DB 108 IVHPOFTYTAQIG--ADIALLELEPVSNNHVTTLPPASETPRPQPCWVTGMDVDN 165

DB 186 GVPLRPQTLQKLVPIIDSEVCSHLVWKGAGGPR---ITEDMTCAGYLEGEBDACLGD 241

DB 166 DERLPFRPLQKQKVVIMENHICDAKYHGLGAYTGDDVRIYRDMILCAG--NTRRDSQGD 223

DB 242 SGGPLKQVDAWLLAGTISWEGCAERNRPQVYISLSAHRSWVEKTV 289

DB 224 SGGPLVCKVNGTWLDAGVSWGEGCAQPNRPQYITRVYLYDMTHHYV 271

RESULT 9

C35863


```

F:26-262/Domain: trypsin homology <TRY>

Query Match          30.3%; Score 522; DB 2; Length 270;
Best Local Similarity 38.0%; Pred. No. 8.5e-37;
Matches 109; Conservative 52; Mismatches 84; Indels 42; Gaps 9;

QY 21 LLLSLTALLNARIRPVPCAGCPQOLNRNVGCGESTDSEMPPIYSIOKNGF--HHGAG 77
    ||||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 LLLALLPLFSLMHRSPFLCQEWG-----YGGGAPGCKMWOVSLRNFETYMRHPCGG 56
    ||||| : : : | : : : : : : : : : : : : : : : : : : : : : : : : :

QY 78 SLTSLRWYITAAHC-----FKDNLNKEPYLESVLLGAWOLGNGPSRSQKVGAWVE 127
    || : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 SLIHQWVLTAAHCQPTIADENKRVQDLRKQYL-----YHDLTAAVSRHII 103
    || : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :

QY 128 PHPV-YSMKEGCAIALVRLERSIQSESVLPICLPASIHLPNTCHWISGWSIODG 186
    || : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 THPTTYATQNG--ADIALEELKPNVNISHHVPSLPPASETFPSGTLCAWYGMGNIND 161
    || : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :

QY 187 VPLPHQTLQKLKVEIILISEVCSHLYMGAGQGP-----ITEDMLCAGYLEGPRACLGDS 242
    ||||| : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 VSLPPPLFLKEQVQVVENOLCDLKYHHGCVTTGDNIHIVRDMCLACGN-EG-HDSCGDS 219
    ||||| : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :

QY 243 GGPLMCQVDGAMLLAGIISWEGCGAERNRPQGVYISLSAHRSMVEKIV 289
    ||||| : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 GGPLVCKVNGTMDLAGVYSWEGCGALPNRPGLYTRVYTYLLDMIHRYV 266
    ||||| : : : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :

Search completed: August 13, 2002, 08:48:00
Job time: 133 sec

```

```
Search completed: August 13, 2002, 08:48:00
Job time: 133 sec
```

Tue Aug 13 16:27:47 2002

us-10-040-803-7.rpx

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:47:37 ; Search time 13.48 Seconds

(without alignments)
910.541 Million cell updates/sec

Title: US-10-040-803-7

Perfect score: 1720
Sequence: 1 MVVSGAPPALGGCLGTFSTS.....AQGGALRAPSQSGAARS 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1720	100.0	317	1 BSS4_HUMAN	O9gzr4 homo sapien
2	1270.5	73.9	306	1 BSS4_MOUSE	O9er10 mus musculu
3	739	43.0	290	1 MPN_HUMAN	O9bqr3 homo sapien
4	607.5	35.3	343	1 PSS8_HUMAN	O1651 homo sapien
5	584	34.0	342	1 PSS8_RAT	O9esr7 rattus norv
6	578.5	33.6	273	1 TRYT_SHEEP	O9xsm2 ovis aries
7	574.5	33.4	342	1 PSS8_MOUSE	O9esd1 mus musculu
8	569	33.1	321	1 TRYG_HUMAN	O9nrr2 homo sapien
9	554	32.2	276	1 MCT6_MOUSE	P21845 mus musculu
10	546	31.7	275	1 TRYT_CANFA	P15944 canis fami
11	538.5	31.3	274	1 MCT6_RAT	P50343 rattus norv
12	532.5	31.0	275	1 TRYT_PIG	O9nrd1 sus scrofa
13	531.5	30.9	314	1 TEST_HUMAN	O9yem0 homo sapien
14	530.5	30.8	275	1 TRB2_HUMAN	P20231 homo sapien
15	530	30.8	324	1 TEST_MOUSE	O9jhj7 mus musculu
16	529.5	30.8	275	1 TRB1_HUMAN	O15661 homo sapien
17	526.5	30.6	275	1 TRYA_HUMAN	P15157 homo sapien
18	522	30.3	270	1 TRYT_MERUN	P50442 metionas un
19	519.5	30.2	273	1 MCT7_RAT	P27435 rattus norv
20	519.5	30.2	311	1 TRYG_MOUSE	O9qul7 mus musculu
21	518	30.1	273	1 MCT7_MOUSE	O02844 mus musculu
22	490.5	28.5	455	1 TMS5_MOUSE	O9er04 mus musculu
23	482.5	28.1	457	1 TMS5_HUMAN	O9h3s3 homo sapien
24	481	28.0	638	1 KAL_HUMAN	P03952 homo sapien
25	475.5	27.6	625	1 KAL_HUMAN	P03951 homo sapien
26	473	27.5	638	1 KAL_MOUSE	P26262 mus musculu
27	472.5	27.5	437	1 TMS4_HUMAN	O9nrs4 homo sapien
28	472.5	27.5	454	1 TMS4_HUMAN	P57727 homo sapien
29	469.5	27.3	638	1 KAL_RAT	P14772 rattus norv
30	463	26.9	416	1 HEPG_MOUSE	O35453 mus musculu
31	462	26.9	492	1 TMS2_HUMAN	O15393 homo sapien
32	460.5	26.8	812	1 PLMN_BOVIN	P06868 bos taurus
33	457.5	26.6	1034	1 ENTK_PIG	P98074 sus scrofa

34	453.5	26.4	490	1 TMS2_MOUSE	O9j1g8 mus musculu
35	452	26.3	417	1 HEPG_HUMAN	P05981 homo sapien
36	451.5	26.2	855	1 STR14_MOUSE	P56677 mus musculu
37	450	26.2	416	1 HEPG_RAT	O05511 rattus norv
38	450	26.2	810	1 PLMN_MACAU	P12545 macaca mula
39	450	25.9	338	1 PLMN_HORSE	P80010 equus caball
40	446	25.9	790	1 PLMN_PIG	P06867 sus scrofa
41	446	25.9	810	1 PLMN_HUMAN	P00747 homo sapien
42	445	25.9	333	1 PLMN_CANFA	P80009 canis fami
43	445	25.9	786	1 STUB_DROME	O05319 drosophila
44	444	25.8	1019	1 ENTK_HUMAN	P98073 homo sapien
45	443.5	25.8	1035	1 ENTK_BOVIN	P98072 bos taurus

ALIGNMENTS

```

RESULT 1
ID      BSS4_HUMAN
AC      O9GZNR4: 043342;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4)
DE      (SP0011A).
GN      PRSS26 OR PRSS22 OR BSSP4.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RX      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RA      Mitsui S., Okui A., Kominami K., Yamaguchi N.;
RT      "Cloning and characterization of a human brain-specific serine
RT      protease, hbssp-4."
RN      Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
[2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Pancreas;
RA      Wong G.W., Stevens R.L.;
RT      "Identification of a new member of the chromosome 16 family of serine
RT      proteases."
RN      Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
[3]
RP      SEQUENCE OF 47-317 FROM N.A.
RA      Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA      Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA      Goodwin L., Bryant J., Tesmer J., Melnick L., Devenire J., White S.,
RA      Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RL      Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
-1- TRYPsin FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL: AB010778; BAB20263.1; -
EMBL: AF321182; AAG35070.1; -
EMBL: AC003965; AAB93671.1; -
DR      MEROPS: S01.252; -
DR      InterPro: IPR001314; Chymotrypsin.
DR      InterPro: IPR001254; Trypsin.
DR      pfam: PF00089; trypsin.1
DR      PRINTS: PR00722; CHYMOTRYPsin.
DR      SMART: SM00020; TRYP_Spc: 1.
DR      PROSITE: PS50240; TRYPsin_DOM: 1.

```


RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ306593; CAC35467.1; -
 DR MEROPS: S01.074; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP_SPC; 1.
 DR PROSITE: PS50240; TRYP_SIN_DOM; 1.
 DR PROSITE: PS00134; TRYP_SIN_HIS; 1.
 DR PROSITE: PS00135; TRYP_SIN_SER; 1.
 KW Hydrolyase; Serine protease; zymogen; signal; glycoprotein.
 FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 34 MARAPST.
 FT CHAIN 35 290 SERINE PROTEASE.
 FT DOMAIN 35 277 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 229 229 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 60 76 BY SIMILARITY.
 FT DISULFID 158 235 BY SIMILARITY.
 FT DISULFID 191 214 BY SIMILARITY.
 FT DISULFID 225 253 BY SIMILARITY.
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 290 AA; 31940 MM; 67BDC93EC70BF7B CRC64;

Query Match 43.0%; Score 739; DB 1; Length 290;
 Best Local Similarity 48.8%; Pred. No. 1.1e-56;
 Matches 139; Conservative 49; Mismatches 87; Indels 10; Gaps 4;

DB 10 LILLCC---FSGQAKAAATACGPRMLRMVGGDTQSGEMPOVSIQKNSHFCGSLI 65
 81 TSSRWYTAHCFKDLNPKYFSLVLLGAMOLGPGSRQKGVANWEPHYVSMKEGACA 140
 66 AAGWVLTAAHCFR-NMSESLKIVLLGAROLVQPGPHMYARVROVESNPLYQ-CTASSA 123
 QY 21 LLLASTALINARIPVPPAGCKPQOLNRYVGGEDSTSEMPVYSIQKNGTHHCAGSL 80
 141 DIALYRLERSIOFSERVLPICLPDASINLPPTHGHWISGWSIQDGVLPHPQTLQKLV 200
 124 DVALVELLEPVPFTNYILPVCLPDSVIFETGMNWMVGMSPSEEDLLPERILQKLV 183
 DB 201 PTIDSEVCHLKWARGOGS---PTEDMLCAGYEGEDACIGSGGGLMCOVGVAMLL 256
 184 PLIDPCKMLLSKOTERYQKTKNDMLCAGFEKGKADCKGSGGPRVLCVQGSMLQ 243
 QY 257 AGIISWEGCAERNRPGVYISLSAHSRWEKIVQGVOLGRRAQ 301
 244 AGVISMEGECARQNRPGVYIRVTAHNMVIRIPLKIQFOPARLGG 288
 DB
 RESULT 4
 PSS8_HUMAN STANDARD; PRT; 343 AA.
 ID PSS8_HUMAN
 AC 01651; OSUCA3;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Prostaasin precursor (EC 3.4.21.-).
 GN PRSS8.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Prostate;
 RX MEDLINE=95286644; PubMed=7768952;
 RA Yu J.X., Chao L., Chao J.;
 RT Molecular cloning, tissue-specific expression, and cellular
 RT localization of human prostaasin mRNA.
 RL J. Biol. Chem. 270:13483-13489(1995).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC TISSUE-Placenta;
 RA Strassberg R.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE OF 45-64.
 RC TISSUE-Semen;
 RX MEDLINE=94308140; PubMed=8034638;
 RA Yu J.X., Chao L., Chao J.;
 RT Prostaasin is a novel human serine protease from seminal fluid.
 RT Purification, tissue distribution, and localization in prostate
 RT gland.
 RL J. Biol. Chem. 269:18843-18848(1994).
 CC -1- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY.
 CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A
 CC DISULFIDE BOND.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF
 CC ITS C-TERMINUS.
 CC -1- TISSUE SPECIFICITY: FOUND IN PROSTATE, LIVER, SALIVARY GLAND,
 CC KIDNEY, LUNG, PANCREAS, COLON, BRONCHUS AND RENAL PROXIMAL TUBULAR
 CC CELLS. IN THE PROSTATE GLAND IT MAY BE SYNTHESIZED IN EPITHELIAL
 CC CELLS, SECRETED INTO THE DUCTS, AND EXCRETED INTO THE SEMINAL
 CC FLUID.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L41351; AAC41759.1; -
 DR EMBL: U33446; AAB19071.1; -
 DR EMBL: BC001462; AAH01462.1; -
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.159; -
 DR MTM: 600823; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP_SPC; 1.
 DR PROSITE: PS50240; TRYP_SIN_DOM; 1.
 DR PROSITE: PS00134; TRYP_SIN_HIS; 1.
 DR PROSITE: PS00135; TRYP_SIN_SER; 1.
 KW Hydrolyase; Serine protease; zymogen; signal; glycoprotein;
 KW Transmembrane.
 FT SIGNAL 1 29 POTENTIAL.
 FT PROPEP 30 32 ACTIVATION PEPTIDE.
 FT CHAIN 33 44 PROSTASIN LIGHT CHAIN.
 FT CHAIN 45 322 PROSTASIN HEAVY CHAIN.
 FT PROPEP 323 343
 FT TRANSMEM 320 340
 FT DOMAIN 45 286 SERINE PROTEASE.
 FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 70 86 BY SIMILARITY.
 FT DISULFID 168 244 BY SIMILARITY.

FT DISULFID 201 223 BY SIMILARITY.
 FT DISULFID 234 262 BY SIMILARITY.
 FT ACT_SITE 85 85 CHARGE RELAY SYSTEM.
 FT ACT_SITE 134 134 CHARGE RELAY SYSTEM.
 FT ACT_SITE 238 238 CHARGE RELAY SYSTEM.
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 343 AA; 36431 MM; 98DD6447F5A8C1B2 CRC64;

Query Match 35.3%; Score 607.5; DB 1; Length 343;
 Best Local Similarity 44.2%; Pred. No. 2.9e-45;
 Matches 129; Conservative 42; Mismatches 108; Indels 13; Gaps 7;

QY 10 LGGGCLGFTSL--LAASTALINARIPVPACGKPOOLNRVGGEDSTDEMPYIS 66
 DB 7 LGPQGLAVALLYLGLFRSGTGAEGAEP---CGVAPGA-RITGSSAVAGQMWQVS 61
 QY 67 IQKNGTHHCAGSLTSRWYITAAHCFKDNINKPYLSVLLGAMQCNPGSRQKGVANV 126
 DB 62 ITTEGVAVCGSLVSEQWVLSAHCPRSEHKE-AVEVKLGALHOLDYSERDAVSTLKD 120
 DB 127 EPHVYSWKEGACADIALVRLERSIOFSERVLPCDPASIHLPNTHGMISGWSIODG 186
 DB 121 IHPHSY-LOHSGOGDIALLOLSRPTFSKIRPICLPANASFPNGLHCTVGMGHVAPS 179
 QY 187 VPLPHQTLQKLVPIIDSEVCSHLYWGA-GQGP--ITEDMLCAGYLEGEPDCLGDSG 243
 DB 180 VSLTFEPQLQLEVPILSRFTCNCLYNIDAKPEPHFQEDMVCAQYVGGKADCGDSG 239
 QY 244 GPLMCGVDGAMLAGIISWEGCAERNRPGVYISLSAHSWVEKYQVQLR 295
 DB 240 GPLSCPEVGLWYLTGIVSWGDAGANRPGVYTLASSVYSKVTLEQPR 291

RESULT 5
 PSS8 RAT STANDARD; PRT; 342 AA.
 AC 09ESB7; 09ER01;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Proctasin precursor (EC 3.4.21.-).
 GN PSS8.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 [1]
 SEQUENCE FROM N.A.
 TISSUE=Kidney;
 Adachi M., Kitamura K., Miyoshi T., Tomita K.;
 Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP SEQUENCE FROM N.A.
 RA Wang C.;
 RT "Molecular cloning and expression of rat proctasin."
 RU Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A
 CC DISULFIDE BOND (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF
 CC ITS C-TERMINUS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AB017638; BAB20281.1; -
 DR EMBL; AF202076; AAG32641.1; -
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 KW Hydrolase; Serine protease; zymogen; signal; glycoprotein;
 KW Transmembrane.
 FT SIGNAL 1 29
 FT PROPEP 30 32 POTENTIAL.
 FT CHAIN 33 44 ACTIVATION PEPTIDE (BY SIMILARITY).
 FT CHAIN 45 44 PROCTASIN LIGHT CHAIN.
 FT PROPEP 323 342 PROCTASIN HEAVY CHAIN.
 FT TRANSMEM 320 340 BY SIMILARITY.
 FT DOMAIN 45 286 POTENTIAL.
 FT DISULFID 37 154 SERINE PROTEASE.
 FT DISULFID 70 86 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 168 244 BY SIMILARITY.
 FT DISULFID 201 223 BY SIMILARITY.
 FT DISULFID 234 262 BY SIMILARITY.
 FT ACT_SITE 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 17 17 I -> V (IN REF. 1).
 FT CONFLICT 292 292 A -> V (IN REF. 1).
 SQ SEQUENCE 342 AA; 36843 MM; 5ED1AF05D9213B98 CRC64;

Query Match 34.0%; Score 584; DB 1; Length 342;
 Best Local Similarity 43.0%; Pred. No. 3e-43;
 Matches 125; Conservative 43; Mismatches 113; Indels 10; Gaps 7;

QY 15 LGFTSLLALSTALINARIPVPACGKPOOLNRVGGEDSTDEMPYISICKNG 71
 DB 9 LGQELAFILLILGLLO-SRIGADGTGASGAVIIO-PRITGGGSAKPGQWQVSITVNG 66
 QY 72 THHCAGSLTSRWYITAAHCFKDNINKPYLSVLLGAMQCNPGSRQKGVANVPEHPV 131
 DB 67 VHVCGGSLVSNQWVLSAHCPRSEHKE-EYVGLGALHOLDSPENDIVYHTVAOIISSS 125
 QY 132 YSMKEGACADIALVRLERSIOFSERVLPCDPASIHLPNTHGMISGWSIODGVP 191
 DB 126 YR-EGSGOGDIALIRLSPTVFSRYRIPCIPANASFPNGLHCTVGMGHVAPVSLQ 184
 QY 192 POTLOKLVPIIDSEVCSHLY-WRGAQGP--ITEDMLCAGYLEGEPDCLGDSGPT 248
 DB 185 PRPLQLEVPILSRFTCNCLYNINAVPEEPHTIOQDMICAGYVGGKADCGDSG 244
 QY 249 QVDCAMLAGIISWEGCAERNRPGVYISLSAHSWVEKYQVQLRGRAQ 299
 DB 245 PIDGLWYLTGIVSWGDAGANRPGVYTLSTVSMWIIHVALDELPRAPVQ 295

RESULT 6
 TRYT_SHEEP STANDARD; PRT; 273 AA.
 AC 09XSM2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trypsin 2 precursor (EC 3.4.21.59).
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID=9940;
 [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Abomasum;
 RA Pemberton A.D., McAllister S.M., Huntley J.F., Mackellar A.,
 RA Collie D.D., McMillan L., Scudamore C.L., Miller H.R.P.;

"DNA sequence of sheep mast cell tryptase and its immunolocalisation in lung, skin and gut in comparison with sheep mast cell proteinase-1.";

Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.

FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION RESPONSE OF THIS CELL TYPE.

CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but with more restricted specificity than trypsin.

SUBUNIT: HOMOTETRAMER (By similarity).

SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON MAST CELL ACTIVATION.

SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1: ALSO KNOWN AS THE TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.

SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Y18224; CABA1989.1; -

HSSP: P20231; 1A0.

MEROPS; S01.143; -

InterPro: IPR001314; Chymotrypsin.

InterPro: IPR001254; Trypsin.

Pfam: PF00089; trypsin; 1.

PRINTS: PR00722; CHYMOTRYPSIN.

SMART: SM0020; TRYPSIN_DOM; 1.

PROSITE: PS00240; TRYPSIN_DOM; 1.

PROSITE: PS00134; TRYPSIN_HIS; 1.

PROSITE: PS00135; TRYPSIN_SER; 1.

KW Hydroxylase; Serine protease; Signal; Glycoprotein; Zymogen.

FT SIGNAL 1 18

FT PROPEP 19 28

FT CHAIN 29 273

FT ACT_SITE 72 72

FT ACT_SITE 119 119

FT ACT_SITE 222 222

FT DISULFID 57 73

FT DISULFID 153 228

FT DISULFID 186 209

FT DISULFID 218 246

FT CARBOHYD 231 231

SEQUENCE 273 AA; 30288 MW; DE9BA79218C3E67D CRC64;

Query Match 33.6%; Score 578.5; DB 1; Length 273;

Best Local Similarity 41.7%; Pred. No. 7.1e-43;

Matches 115; Conservative 53; Mismatches 93; Indels 15; Gaps 7;

21 LLLASTAIIINARIIPVPACGKPOQLNRVVGSDSDSEMPWIVSIQ---KNGTHHCAG 77

1 LHLTALLLSL--VSAAPAGCALORSIGIKREKAGSPWQVSLRVDDQYWRHCGG 59

2 LHLTALLLSL--VSAAPAGCALORSIGIKREKAGSPWQVSLRVDDQYWRHCGG 59

78 SLTSRWITTAHCKFNKLPYLFVILGAMQNGSGSOKKGVAMVEHPHYSKKEG 137

60 SLIHQWVLAHCTIGPELOSPDFVOLRQHL---YDQRLPIRATVIRPHRYAVEN 116

138 ACADIALVRLERSIQSERVLPICLPASIHLPNTHCWISGWSIODGVLPHPQTLQK 197

117 G-ADIALQLQLEPYSISRHQVPLTPRASEFPESQCVWGTGMDVDVNGRPLRPYELKQ 175

198 LKVPIDSEVCSHLYWRAGOG---PIT-EDMLCAGYLEGRDCLDSDSGPLMCQVDA 253

176 VKVPVENSVCMDKWHSLSTDYSPVIVQEDNLCAG--DGGRDSCQSDSGSLPCKVNGT 233

254 WLAGIISWGEGCAERNRPVYISLSAHSRVEKIV 289

234 WLGAGVSWMGCGAKRNPFGIYITNTISLDMIHQYV 269

RESULT 7

PS58_MOUSE STANDARD; PRT; 342 AA.

AC Q9ESD1;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Prostatin precursor (EC 3.4.21.-) (Channel activating protease 1).

GN PS58 OR CAP1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxId=10090;

[1]

SEQUENCE FROM N.A.

RA Vaugniat G., Vallet V., Fowler-Jaeger N., Bens M., Farman N.,

RA Courtois-Couty N., Vandewalle A., Rossier B.C., Hummel E.;

RT "Activation of the amiloride-sensitive sodium channel by the mouse

RT serine protease mCAP1 expressed in a principal kidney cell line."

RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY (BY

CC SIMILARITY). ACTIVATES AMILORIDE-SENSITIVE SODIUM CHANNELS.

CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A

CC DISULFIDE BOND (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF

CC ITS C-TERMINUS (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

CC FRAMESHIFT IN POSITION 339.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL: AF188613; AAG17054.1; ALT_FRAME.

DR InterPro: IPR001254; Trypsin.

DR Pfam: PF00089; trypsin; 1.

DR SMART: SM0020; TRYPSIN_DOM; 1.

DR PROSITE: PS00240; TRYPSIN_DOM; 1.

KW Hydroxylase; Serine protease; Zymogen; Signal; Glycoprotein;

KW Transmembrane.

FT SIGNAL 1 29

FT PROPEP 30 32

FT CHAIN 33 44

FT CHAIN 45 322

FT PROPEP 323 342

FT TRANSMEM 320 340

FT DOMAIN 45 286

FT DISULFID 37 154

FT DISULFID 70 86

FT DISULFID 168 244

FT DISULFID 201 223

FT DISULFID 234 262

FT ACT_SITE 85 85

FT ACT_SITE 134 134

FT ACT_SITE 238 238

FT CARBOHYD 110 110

FT CARBOHYD 159 159

SEQUENCE 342 AA; 36729 MW; 0620DE88ED187D0F CRC64;

Query Match 33.4%; Score 574.5; DB 1; Length 342;

Best Local Similarity 42.0%; Pred. No. 2e-42;

Matches 124; Conservative 46; Mismatches 110; Indels 11; Gaps 8;

10 LGGCGTGTFTSLIILASTAIINARIIPVPACGKPOQLNRVVGSDSDSEMPWIVSIQ 68

Db 7 LGAGLEAVTILLLLG--LLQSGIRADGTEASGAVIQ--PRITGGSGANPGQWPMQVST 63
 QY 69 KNGTHAGCSILTSRWYTAHCF-KDNLNKPYLFSVLGAMOLGNGSRQKGVAYME 127
 Db 64 YDGNHGGSSLVNKNWVAHCFPREHSEAL--EVLGAHGLDSYSDNYVHTYVQIT 121
 QY 128 PPHVYSKKEGACADIALVRLERSIQFSERVLPICLPDASTHLPNTNCHWISGWSIODGV 187
 Db 122 THSSYR-EEGSGQDIAPFRLSSVFTSRVIRPCLPAANASFPNGLHCVTGMGHVAPSV 180
 QY 188 PLHPOTLQKLPILIDSEVCSHLV-WRGAQGP--ITEDMLCAGLIEBFRACLDSSG 244
 Db 181 SLDPRLPRLQLEVLISRESCVSLNINAVPEEPHTIQDMICAGYVKGKDCQDSSG 240
 QY 245 PLMCOVDGAMVLGILISWEGCAERNRPGVYISLSAHSRWEKTVQGVOLR 295
 Db 241 PLSCPMEGIWYLAGIYVWGDAGCAPRNPVYITLTSTYASMHVHVALQPR 291

QY 8
 HUMAN
 TRYG-HUMAN STANDARD; PRT; 321 AA.
 Q9NR02: Q9NR08: Q9C015: Q9DBR2:
 16-OCT-2001 (Rel. 40, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trypsin gamma precursor (EC 3.4.21.-) (Transmembrane trypsinase).
 GN TP5G1 OR TMT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2).
 RA MEDLINE-20302813; PubMed-10843716;
 RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaro M.,
 RA Wolters P.J., Verghese G.M.:
 RT "Characterization of human gamma-trypsins, novel members of the
 RT chromosome 16p mast cell trypsin and proctasin gene families.";
 RL J. Immunol. 164:6566-6575(2000).
 RN (2)
 RP SEQUENCE FROM N.A.
 RA MEDLINE-99452974; PubMed-10521469;
 RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C.,
 RA Friend D.S., Krilis S.A., Stevens R.L.:
 RT "Identification of a new member of the trypsin family of mouse and
 RT human mast cell proteases which possesses a novel COOH-terminal
 RT hydrophobic extension.";
 RL J. Biol. Chem. 274:30784-30793(1999).
 RN (3)
 RP SEQUENCE OF 220-321 FROM N.A.
 RA Miltman S., Agnew W.S.:
 RT "Organization and alternative splicing of CACNA1H.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ANCHORED (potential).
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues.
 CC -1- POLYMORPHISM: There are two alleles; gamma-I and gamma-II which
 CC differ by 5 residues.
 CC -1- SIMILARITY: BELONGS TO TRYPSIN FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. TRYPSIN SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; AF191031; AAF76457.1; -
 CC DR EMBL; AF195508; AAF76458.1; -
 CC DR EMBL; AF175759; AAF03697.1; -
 CC DR EMBL; AF175522; AAF03695.1; -

DR EMBL; AF223563; AAC48852.2; -.
 DR MEROPS; S01.028; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP-SPC; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
 KM Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;
 KM Transmembrane; Polymorphism.
 FT SIGNAL 1 19
 FT CHAIN 20 36
 FT CHAIN 38 321
 FT TRANSMEM 284 304
 FT ACT_SITE 78 78
 FT ACT_SITE 125 125
 FT ACT_SITE 222 222
 FT DISULFID 26 145
 FT DISULFID 63 79
 FT DISULFID 159 228
 FT DISULFID 192 210
 FT DISULFID 218 246
 FT CARBOHYD 85 85
 FT VARIANT 60 60
 FT VARIANT 126 126
 FT VARIANT 132 132
 FT VARIANT 204 204
 FT VARIANT 288 288
 FT CONFLICT 160 160
 FT SEQUENCE 321 AA; 33827 MW; FF7B06B3C4A962D CRC64;

Query Match 33.1%; Score 569; DB 1; Length 321;
 Best Local Similarity 41.1%; Pred. No. 5.6e-42;
 Matches 123; Conservative 44; Mismatches 108; Indels 24; Gaps 8;

QY 15 LGFTTSLLLASTAIALNARIIPVPACGKPOQLN--RVVGGEDSDNSEMPWIVSIQNG 71
 Db 3 LGAGCILLT---AVPVSLRTLOPGCGRPVSDAGRIYVGHNAAPGAMPQASLRIR 59
 QY 72 THHAGSILTSRWYTTAHCFFKDLNKPFLPSVLGAMOLGNGSRQKGVAYMEPHV 131
 Db 60 MHVCGSSILSPQWVLTAAHCFSGSLNSSD--YQVHLGELTETLSPHST--VROIIHSS 115
 QY 132 YSNREGACADIALVRLERSIQFSERVLPICLPDASTHLPNTNCHWISGWSIODGVPLP 191
 Db 116 PSQPGTSGDIALVELSVPTLSRLIPVCLPEASDFGICGICWYTGWYTRGEPLP 175
 QY 192 PQTQKLPILIDSEVCSHLVWRGAQGPITEDMLCAGLIEBFRACLDSSGGLMCOVD 251
 Db 176 PYSLREKVASVDETCRRDY-PGPGGSILOPMLCA---RGPGDACQDSDSGPLVCQVA 231
 QY 252 GAWLTLGILISWEGCAERNRPGVYISLSAHSRWEKTVQGVOLRGAQGS---GALRAP 307
 Db 232 GAWVQAGIVSMGEGCAPRNPVYITRVPAYVNMIRRI-----TASGSESSEGYRLP 283

RESULT 9
 MCT6_MOUSE STANDARD; PRT; 276 AA.
 AC P21845; Q61962;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mast cell protease 6 precursor (EC 3.4.21.59) (MCP-6) (Trypsinase).
 GN MCT6.

KW Alternative splicing.
 FT SIGNAL 1 21
 FT PROPER 22 31
 FT CHAIN 32 276
 FT ACT_SITE 75 75
 FT ACT_SITE 122 122
 FT ACT_SITE 122 122
 FT ACT_SITE 225 225
 FT DISULFID 60 76
 FT DISULFID 156 231
 FT DISULFID 189 212
 FT DISULFID 221 249
 FT CARBOHYD 133 133
 FT CARBOHYD 230 230
 FT VARSPLIC 221 276
 FT VARSPLIC 221 276
 SQ SEQUENCE 276 AA; 30927 MW; 525B2C9A04A72200 CRC64;

Query Match 32.2%; Score 554; DB 1; Length 276;
 Best Local Similarity 41.6%; Pred. No. 9, 4e-41;
 Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9

QY 21 LLLASTALLNARIYVP-PACGKPOQLNRVGGEDSTSEWPWISIQ--KNGTHICA 76
 Db 6 LLLMLSLDLSLIVSAPPRA---NORVGIYGHGSESKPMQVSLRPKLNYMHIFGC 61
 QY 77 GSILTRWVITFAHCKNDLNKPYLSVLGAQL--GNPGSISQKVGAVMEPHVYISQ 134
 Db 62 GSILHQQWVLTAAHCYGPPIKSPQLRVLQREQYLYGD---QLSLNRIYVHPHYT 116
 QY 135 KKGACADIALVRLERSIOFSERYLPICLDASITLPPNTHCWSIGWSIODGVLPLHPOT 194
 Db 117 AEGG-ADVALLLELVENVNSTHHPISLPAPSETPPGTSCHWTGMDINDPELRPPRP 175
 QY 195 LQKLVPIIDSEVCSHLVYRGAGOG---PITD-MCAGYLEGERACLDGSGPLMCOY 250
 Db 176 LKQVVPYIVENSICDRKTYTGLTGDPPIYHDGMICAG--NTRRSCGDSGGLVCRV 233
 QY 251 DGAMLLAGIISWEGCAERNRPQVYISLSAHSWEKIV 289
 Db 234 KGTWLQAGVSWEGCAQCPNKPGIYTRVYIYDMIHRYV 272

RESULT 10
 TRY1_CANFA STANDARD; PRT; 275 AA.
 ID TRY1_CANFA
 AC P15944;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trypsin precursor (EC 3.4.21.59).
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxId=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89352460; PubMed=2504277;
 RA Vanderslice P., Craik C.S., Nadel J.A., Caughey G.H.;
 RT "Molecular cloning of dog mast cell trypsinase and a related protease:
 structural evidence of a unique mode of serine protease activation.";
 RT Biochemistry 28:4148-4155(1989).
 CC -1- FUNCTION: TRYPSIN IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST
 CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION
 CC RESPONSE OF THIS CELL TYPE.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but
 CC with more restricted specificity than trypsin.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON
 CC MAST CELL ACTIVATION.
 CC -1- SIMILARITY: BELONGS TO TRYPSIN-LIKE TRYPSIN-LIKE TRYPSIN-LIKE TRYPSIN-LIKE
 CC TRYPSIN FAMILY. STRONGEST TO OTHER TRYPSINS.
 CC
 CC This SWISS-PROT entry is copyright © 1994 by the Swiss Institute of Bioinformatics.

DR EMBL; M24664; AAA308
DR PIR; A32410; A32410.

DR HSSP; P20231; JAAO.
DR MEBOPE; S01 143: -

DR MEMOFS, 501.143; ;
DR InterPro; IPR001314; Chymotrypsin.
DR

DR InterPro; IPR001254; Trypsin.
DR Pfam: PF00089.4; trypsin.1

DR PLAIN, CYPRESS, I.
DR PRINTS; PR00722; CHYMOTRYPSIN.

```
DR SMART; SM00020; Tryp_Spc; 1.  
DE PROCTITE; PS50240; TRYPSIN DOM; 1
```

```
DR PROSITE; PS00240; TRYPSIN_DOM; 1
DR PROSITE; PS00134; TRYPSIN_HIS; 1
```

DR PROSITE; PS00135; TRYPSIN_SER; 1.
hydrolase: Serine protease: Signal: G]ycoprotein: Zymogen

hyaluronase, 20
 1 SIGNAL POTENTIAL.

PROPEP	21	30	ACTIVATION PEPTIDE (BY SIMILARITY)
CHAIN	31	275	TRYPTASE.

CHAIN	212	CHARGE RELAY SYSTEM (BY SIMILARITY)
FT	74	
ACT_SITE	74	

FT	ACT_SITE	121	121	CHARGE RELAY SYSTEM (BY SIMILARITY
ET	ACT_SITE	224	224	CHARGE RELAY SYSTEM (BY SIMILARITY

FT	DISULFID	59	75	227	227	BY SIMILARITY.
FT	ACI-SIDE	227	227	227	227	BY SIMILARITY.
FT	ACI-SIDE	227	227	227	227	BY SIMILARITY.

FT	DISULEID	155	230	BY SIMILARITY.
ET	DISUREID	188	211	BY SIMILARITY.

FT	DISULFID	220	248	BY SIMILARITY.
FT	DISULFID	220	248	BY SIMILARITY.
FT	DISULFID	220	248	BY SIMILARITY.

FT	132	132	N-LINKED (GLCNAC. . .) (POTENTIAL)
CARBOHYD	275	AA:	30088 MW: C3B869251F248D5B CRC64;
SEQUENCE			

6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529

Query Match 31.78; Score 546; DB 1; Length 275;

Best Local Similarity 40.98; Pred. No. 4.6e-40;

Matches	112;	Conservative	46;	Mismatches	102;	Indels	14;
---------	------	--------------	-----	------------	------	--------	-----

QY 23 LASTAILNARIPVPACGKPOLNRVGGEDSTDSEMPWISIQKNG---THHCAG

5 T V I A T A I A I G S L - V P V S P A R G A L O R V G I V G C R E A P S K M P W Q V S L R L K G Q Y W R H I C G G

1. The first step is to identify the problem or goal. This involves understanding the current situation and what needs to be achieved.

80 LTRWVITAHCFKDNLNKPYLFSVLLGAWQLGNPGSKSQA VGVAMVEHFVISMNES

Db 64 IHPQWLTAAHCYGFNVCCPEEIRVQLRQHL--YYQDHLIPVNRIVMHPNYYTPEN

140 ADATVRIFRSIOFSERYLPICLPDASIHLPNTHCWSGWSIODGVPLPHPQTLOK

170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

120 ADIATLELEDPVNSAHQPVLTLPRALQIFPIGTPCMVIGMGDVHSGIFEPFFFEENQ

200 VPIDSEVCSHLWYRGAGGP---ITEDMLCAGYLEGERDACLGDSGGPLMCQVDGA

180 VPVNSMCDVOYHGLSTGDGVRIREDMLCAG--NSKSDSCQGDSSGPLYCVRVGRV

[illegible]

QY 256 LAGIISWGECAERNRPGVILSLAHRSWENILV 209

Db 238 QAGVSWGEGCAQPNRPGITYRVAYLLDWHQYV 271

[illegible]

Search completed: August 13, 2002, 06:21:26
Job time: 231 sec

● ● ● ●

1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:47:22 ; Search time 30.21 Seconds
(without alignments)
1815.273 Million cell updates/sec

Title: US-10-040-803-7

Perfect score: 1720

Sequence: 1 MVSGAPALGGGCLGFTS.....AQGGALRAPSGSGAARS 317

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 562222

Database : SPTREMBL_19:*

1: sp.archaea:*

2: sp.bacteria:*

3: sp.fungi:*

4: sp.human:*

5: sp.invertebrate:*

6: sp.mammal:*

7: sp.mhc:*

8: sp.organelle:*

9: sp.phage:*

10: sp.plant:*

11: sp.rodent:*

12: sp.virus:*

13: sp.vertibrate:*

14: sp.unclassified:*

15: sp.virus:*

16: sp.bacteriap:*

17: sp.archive:*

Post-processing: Minimum Match 0%

Maximum DB seg length: 200000000

Maximum DB seg length: 200000000

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	1213.5	70.6	297	11	088781	088781 rattus rattu
2	660.5	38.4	389	13	09PVX7	09PVX7 xenopus lae
3	603	35.1	317	13	09DGR3	09dgr3 xenopus lae
4	578.5	33.6	339	11	09JL44	09JL44 mus musculus
5	571	33.2	310	11	09JXC4	09JXC4 mus musculus
6	570	33.1	321	4	096R28	096R28 homo sapien
7	564	32.8	310	11	09OY29	09OY29 mus musculus
8	559.5	32.5	273	6	09XSM1	09XSM1 ovis aries
9	537	31.2	799	11	09DB10	09db10 mus musculus
10	536	31.2	237	6	029464	029464 bos taurus
11	529.5	30.8	329	13	042272	042272 xenopus lae
12	524	30.5	273	11	09JL44	09JL44 mus musculus
13	517.5	30.1	275	4	096R26	096R26 homo sapien
14	491	28.5	806	6	018783	018783 macropus eu
15	485.5	28.2	279	11	099MS4	099ms4 mus musculus
16	480	27.9	282	11	09D413	09D413 mus musculus

ALIGNMENTS

17	480	27.9	322	11	092052	092052 mus musculus
18	475	27.6	624	11	091Y47	091Y47 mus musculus
19	473	27.5	624	11	09D413	09D413 mus musculus
20	472.5	27.5	405	4	096E86	096E86 homo sapien
21	469	27.3	624	6	095ME7	095ME7 oryctolagus
22	463.5	26.9	267	5	09BK47	09BK47 ludia foli
23	462.5	26.9	643	6	097506	097506 sus scrofa
24	462	26.9	492	4	096T73	096T73 homo sapien
25	455.5	26.5	812	11	09R0W3	09R0W3 rattus norv
26	452	26.3	454	6	046506	046506 papio hamad
27	450	26.2	334	6	046507	046507 papio hamad
28	445.5	25.9	812	11	091WJ5	091WJ5 mus musculus
29	445.5	25.9	1524	13	091674	091674 xenopus lae
30	445	25.9	787	5	09VEY6	09VEY6 drosophila
31	445	25.9	810	4	015146	015146 homo sapien
32	439.5	25.6	490	11	0920K3	0920K3 rattus norv
33	437.5	25.4	367	11	070169	070169 mus musculus
34	435.5	25.3	581	4	09BYE2	09BYE2 homo sapien
35	435	25.3	263	11	09DC86	09DC86 mus musculus
36	435	25.3	277	5	096899	096899 scolopendra
37	434	25.2	265	5	017800	017800 caenorhabdi
38	433	25.2	329	6	09GL10	09GL10 ovis aries
39	433	25.2	537	4	09BYE1	09BYE1 homo sapien
40	432.5	25.1	767	13	09DGR2	09DGR2 xenopus lae
41	430.5	25.0	366	11	070170	070170 mus musculus
42	430	25.0	263	11	09CR35	09CR35 mus musculus
43	429.5	25.0	261	13	09W7Q4	09W7Q4 paracichthys
44	429.5	25.0	279	11	09QZ74	09QZ74 rattus norv
45	428.5	24.9	249	13	09W6X0	09W6X0 notothenia

RESULT 1						
088781						
ID	088781	PRELIMINARY:	PRT:	297	AA.	
AC	088781:					
DT	01-NOV-1998 (TREMBLrel. 08, Created)					
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)					
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)					
DE	SERINE PROTEASE PRECURSOR (FRAGMENT).					
GN	BSP2.					
OS	Rattus rattus (Black rat).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
OX	NCBI_TaxID=10117;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-FISHER: TRISUB-BRAIN;					
RX	MEDLINE=98389725; PubMed=9722524;					
RA	Davies B.D., Pickard B.S., Steel M., Morris R.G., Lathe R.;					
RT	J. Biol. Chem. 273:23004-23011(1998).					
RL	"-i" SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE					
CC	TRYPSIN FAMILY.					
DR	EMBL: AJ005642; CA06644.1; -					
DR	HSSP: P00763; IDPO.					
DR	MEROPS: S01.252; -					
DR	InterPro: IPR001314; Chymotrypsin.					
DR	InterPro: IPR001254; Trypsin.					
DR	Pfam: PF00089; Trypsin. 1.					
DR	PRINTS: PR00722; CHYMOTRYPSIN.					
DR	SMART: SM00020; TRYP-Spe; 1.					
DR	PROSITE: PS00240; TRYPSIN_DOM; 1.					
DR	PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.					
DR	PROSITE: PS00135; TRYPSIN_SER; 1.					
KW	Hydrolase; Protease; Serine protease; Signal.					
FT	NON_TER	1				
FT	SIGNAL	<1	23		POTENTIAL.	
FT	CHAIN	24	297		SERINE_PROTEASE.	
SO	SEQUENCE	297	AA;	32086	MM;	6DA25C633D6AB55 CRC64;

Query Match 70.6%; Score 1213.5; DB 11; Length 297;
 Best Local Similarity 74.7%; Pred. No. 3e-104; Indels 9; Gaps 1;
 Matches 222; Conservative 26; Mismatches 40; Indels 9; Gaps 1;

QY 6 APALGGGCGTFTSLLLASTAILNARIYPVPCGKPOQLNRVVGEDSDSEMPWIV 65
 DB 5 SPGL-----TFILLPSATVSANIRSGDKPOQLNRVVGEDSDADQMPWIV 55
 QY 66 SIOKNGTHHCAGSLTSRWVITAAHCFKDNLPYFVLGAMOLGNGSRQKGVAM 125
 DB 56 SIKNGSHHCAGSLTNRWVSAHCFSSNMDKPSPYVLGAMKLGNGSPRQKGLAS 115
 QY 126 VEPHRYVSMKGCADIALVRLERSIOFSERVLPICLPDASIHLPNTCMISGMSTOD 185
 DB 116 VLPHPRKSRKEGHADIALVRLERPIQFSRIIPICLPDSSVHLPPNTCMWAGMSIOD 175
 QY 186 GVPPLPHTLOKLVPIIDSEVCSHLWRGAGGPTTDMLCAGYLEGERDACLGDGSGP 245
 DB 176 GVPPLPHTLOKLVPIIDSEVCSHLWRGAGGPTTDMLCAGYLEGERDACLGDGSGP 235
 QY 246 LMCQVDCAMLLAGIISWEGCAERNRPGVYISLHRSWVEKIVQVQLRGAQGGG 302
 DB 236 LMCQVDHMLLTGIIISWEGCAERNRPGVYISLHRSWVEKIVQVQLRGLADSG 292

RESULT 2
 Q9PVX7 PRELIMINARY; PRT; 389 AA.
 ID Q9PVX7;
 AC Q9PVX7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PEPTIDASE SPECIFIC SERINE PROTEASE.
 GN XEPSTIN.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K.;
 RT "The expression control of xepsin by non-axial and planar
 RT posttericizing signals in xenopus epidermis.";
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 EMBL: AB018694; BAB84941.1; -;
 HSSP: P00763; IDPO.
 MEROPS: S01.048; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP_SPC.1.
 DR PROSITE: PS50240; TRYPsin_DOM.1.
 DR PROSITE: PS00134; TRYPsin_HIS.1.
 DR PROSITE: PS00135; TRYPsin_SER.1.
 KW Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 389 AA; 42375 MW; B31FB4A2F5D1FE63 CRC64;

Query Match 38.4%; Score 660.5; DB 13; Length 389;
 Best Local Similarity 47.0%; Pred. No. 6.2e-53;
 Matches 131; Conservative 42; Mismatches 95; Indels 11; Gaps 6;

QY 40 ACGRPQOLNRVVGEDSDSEMPWIVSIQKNGTHHCAGSLTSRWVITAAHCFKDNLPK 99
 DB 16 ACGVPIVSRIVGMDSKRGEMWQISLSYKSDSGLTDSWWTAAHCT-DSIDVS 74
 QY 100 YLSEVLGAMOLGNGSRQKGVAMVEPHRYVSMKGCADIALVRLERSIOFSERVLP 159
 DB 236 LMCQVDHMLLTGIIISWEGCAERNRPGVYISLHRSWVEKIVQVQLRGLADSG 292

DB 75 Y-YTVYLGAAYOLAPDNSTVSRGVKSTIKHPDFQY-EGSSGDIALIELEKPVFTPIILP 132
 QY 160 ICLPDAIHLPPNTHCWISGWSIODGVLPHTLOKLVPIIDSEVCSHLWRGAGG 219
 DB 133 ICLPDAIHLPPNTHCWISGWSIODGVLPHTLOKLVPIIDSEVCSHLWRGAGG 219
 QY 220 P-----ITDMLCAGYLEGERDACLGDGSGPLMCOVDGAMLLAGIISWEGCAERNRPGV 275
 DB 193 PDFSFTIQEDMVCAGYKGRIDACQSGSGPLVCVNNNNWITOLGIYSWYGCAEPNRPVY 252
 QY 276 ISLSHRSWVEKIVQVQLRGAQGGALRRAPSGSGAA 314
 DB 233 TKVQYQDMWK--TNVPLIVSEGPSV-APSIGPSIA 287

RESULT 3
 Q9DGR3 PRELIMINARY; PRT; 317 AA.
 ID Q9DGR3;
 AC Q9DGR3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE EMERYONIC SERINE PROTEASE-1.
 GN XESP-1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-20363741; PubMed-10903452;
 RA Yamada K.; Takabatake T.; Takeshima K.;
 RT "Isolation and characterization of three novel serine protease genes
 RT from xenopus laevis.";
 RL Gene 252;209-216(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 EMBL: AB038496; BAB08216.1; -;
 HSSP: P00763; IDPO.
 DR MEROPS: S01.048; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP_SPC.1.
 DR PROSITE: PS50240; TRYPsin_DOM.1.
 DR PROSITE: PS00134; TRYPsin_HIS.1.
 DR PROSITE: PS00135; TRYPsin_SER.1.
 KW Hydrolyase; Protease; Serine protease.
 SQ SEQUENCE 317 AA; 34413 MW; EEC78A9F46D138FE CRC64;

Query Match 35.1%; Score 603; DB 13; Length 317;
 Best Local Similarity 40.4%; Pred. No. 1e-47;
 Matches 116; Conservative 52; Mismatches 103; Indels 16; Gaps 5;

QY 10 LGGGCGTFTSLLLASTAILNARIYPVPCGKPOQLNRVVGEDSDSEMPWIVSIQK 69
 DB 1 MCKWLLYVTTLLFVSPHPSISNITTAAPILGSGVFSRIVGDTQAGAMPQVSLF 60
 QY 70 NCTHHCAGSLTSRWVITAAHCFKDNLPYFVLGAMOL-GNPSRSQKGVAMVE 127
 DB 61 NSHICGGSIISDOWILATFHTCI-EHPDLPSCGVRALAYLKNPHEMTKVDIIYIN 119
 QY 128 PH---PVYSMKGCADIALVRLERSIOFSERVLPICLPDASIHLPNTCMISGMSTOD 184
 DB 120 SPENP-----GTSDIALLLKSSPIKFTYILPILCLPASPVPSSGTEDMITGWGTG 173
 QY 185 DGVPPLPHTLOKLVPIIDSEVCSHLWRGAGGPTTDMLCAGYLEGERDACLGDGSGP 240
 DB 174 SEVPLQYPATLOKLVVPIINRDCERKNHINSIVSETELLIOSDQICAGYAGKDGCG 233
 QY 241 DSGPLMCOVDGAMLLAGIISWEGCAERNRPGVYISLHRSWVEK 287

DB 234 DSGGLVCKIQGFVWQAGIVSGMGRCAKNNRGVYTFVPAVPTWSE 280

RESULT 4
ID 099144 PRELIMINARY; PRT; 339 AA.

AC 099144: PRELIMINARY; PRT; 339 AA.
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SIMILAR TO PROTEASE, SERINE, 8 (PROSTASIN).

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;

SEQUENCE FROM N.A.
Strausberg R.;
Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

STRAIN=129/SVJ, AND SWISS; TISSUE=LUNG;

Verghese G.M., Caughey G.H.;

"Molecular cloning and characterization of mouse prostasin, a type I
membrane-associated serine protease of the gamma-tryptase/prostasin
gene family."

Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
TRYPSIN FAMILY.

EMBL: BC003851; AA03851.1; -;
EMBL: AF378086; AAL06320.1; -;
EMBL: AF378085; AAL06319.1; -;

HSP: P00734; IUVS.

MEROPS: S01.158; -;
InterPro: IPR001314; Chymotrypsin.

InterPro: IPR001254; Trypsin.

Pfam: PF00089; trypsin; 1.

PRINTS: PR00722; CHYMOTRYPSIN.

SMART: SM00240; TRYP-Spc; 1.

PROSITE: PS00240; TRYP-SIN_DOM; 1.
PROSITE: PS00135; TRYP-SIN_SER; 1.

Hydrolase; Protease; Serine protease.

SEQUENCE 339 AA; 36216 MW; BC2DE88BC057AF10 CRC64;

Query Match 33.6%; Score 578.5; DB 11; Length 339;
Best Local Similarity 43.0%; Pred. No. 2e-45; Mismatches 109; Indels 11; Gaps 8;

atches 125; Conservative 46; Mismatches 109; Indels 11; Gaps 8;

10 LGGGCGTFTSLLAATAILNAARIP-VPAAGRPQOLNRVGGEDSTDEMPWISIQ 68

7 LGIGLEAVTILLIG-LLQSGIRADTEASCAGVIG-PRITGGSAKPGQMPQVYSIT 63

69 KNGTHHCASLTSRVVITAAHCF-KDINKPYLSVLGAMQOLNPGSRKQVAVWVE 127

64 YGNHVCGLSVSNMVMVSAACFPRHSREAY-EVKIGAHQIDSYSDTYVVAQI 121

128 PHVYSWKGACADIALVLETSIOFSERVLPICLPDASIHLPMTWHISGWSIQGY 187

122 THSSSTR-EGSGGDIALLISPVTFSTRIRICLPANASFPNGLCTVIGWGHVADSV 180

188 PLPHQTLQKLVPIIDSEVCSHLY-WRQAGGP--ITEDMLCAGYLGEGRDACLGDGSG 244

181 SLQTPRLQQLAVPLISRTCSCLYNINAVPEPHTIOODMLCAGYVVGKACGDSGG 240

245 PLMCOVDGAMLLAGITISWEGCAERNRGVYISLSAHSRVKTYQVQLR 295

241 PLSCPEGIWYLAGIVSWGDACGAPNRPGVYTLTSTYASWTHHVAELQPR 291

RESULT 5
ID 091XC4 PRELIMINARY; PRT; 310 AA.

AC 091XC4: PRELIMINARY; PRT; 310 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SIMILAR TO DISTAL INTESTINAL SERINE PROTEASE.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;

SEQUENCE FROM N.A.
Strausberg R.;
Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

EMBL: BC010970; AAH10970.1; -;
Protease.

SEQUENCE 310 AA; 33707 MW; FA126747DEAE0AB6 CRC64;

Query Match 33.2%; Score 571; DB 11; Length 310;
Best Local Similarity 42.4%; Pred. No. 8.9e-45; Mismatches 117; Conservative 50; Mismatches 93; Indels 16; Gaps 9;

atches 117; Conservative 50; Mismatches 93; Indels 16; Gaps 9;

21 LLLASTAILNAARIPV-PPACGKRPQOLNRVGGEDSTDEMPWIS--IQKNGTHHCAG 77

10 LLLL--QILFRAGDILPSVCGHSRDAGKIVGGQDALEQMPQVSLWITEDG-HICGG 65

78 SLTSRWVITAAHCFKLNKPYLFVSLGAMQOLNPGSRKQVAVWVEPHYSWKEG 137

66 SLIEHWVITAAHCFKRSLSN-PSFYHVKVGGITLSLEPHSTLYAVANNITVHPYLMADA 124

138 ACADIALVLETSIOFSERVLPICLPDASIHLPMTWHISGWSIQGYPLPHQTLQK 197

125 SSGDIALVQDPTLRPSPQ-FTPVCLPAAQPLRPGYTCWVTGWCATGE--RDMAVYLQ 180

198 LKVPILIDSEVCSHLY-WRG--AGQPTEDMLCAGYLGEGRDACLDSGGLMCOYDGA 253

181 LAVPLIDSEDEKMYHVGSSLSGERITQSDMLCAGYVEGQKSCQDSGGPLVCSINSS 240

254 WLAGIISWEGCAERNRGVYISLSAHSRVKTYQVQLR 289

241 WTQVGTISWGTGCAAPRPGVYTLTSTYASWTHHVAELQPR 276

RESULT 6
ID 096R28 PRELIMINARY; PRT; 321 AA.

AC 096R28: PRELIMINARY; PRT; 321 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HS TRANSMEMBRANE TRYPTASE, GENE NAME TMT, AF175522.1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_Taxid=9606;

SEQUENCE FROM N.A.

MEDLINE=21096910; PubMed=11157797;

Daniels R.J., Belden J.F., Lloyd C., Horsley S.W., Clark K.,
Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
Higgs D.R.;

"Sequence, structure and pathology of the fully annotated terminal 2
Mb of the short arm of human chromosome 16."

Hum. Mol. Genet. 10:339-352(2001).

EMBL: AE006465; AAK61269.1; -;
Transmembrane.
SEQUENCE 321 AA; 33829 MW; FFF5089EDCAFC73D CRC64;

Query Match 33.1%; Score 570; DB 4; Length 321;
 Best Local Similarity 41.1%; Pred. No. 1.2e-44;
 Matches 123; Conservative 44; Mismatches 108; Indels 24; Gaps 8;

15 LGFTSLTLLASTALINARIIPVPPACKPOOLN---RVVGGEDSTDEMPWIVSIQNG 71
 3 LGACGLLLTLL---AVPGVSLRLTIGCGCPQVSDAGRIYCGHAAPAGAMPQASILRLR 59
 72 THHCGSLTSSRWITTAHCFKDNKRYLFSVLLGAMOLGNPGRSKQVAVWVERHPV 131
 60 VHVCGSLSPQWVLTAAHCFSGSLNSSD-YOVHGLELITLSPHST---VROLTHSS 115
 132 YSMKGCACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWSIGMSIODGVPLPH 191
 116 PGGGPTSGDIALVELSPVLTSSRLPYCLPEASDDCPGRCVMTGRTREEPEPLP 175
 192 PQTLOKLVPIIDSEVCSHLVWAGAGGPTEDMLCAGYLEGERDACLDSGGPLMCOVD 251
 176 PYSLEEVKVSVDTECTCRDXY-PGPGGSITLOPMLCA---KPGPDACODSDSGPLVCOVN 231
 252 GAWLAGITISNCGCAERNRPVYISLSAHSRWKIVQGVOLRRAGG---GALRAP 307
 232 GAWVAGIVSWGCGGRPNRPVYTRPVYVNMIRRH1-----TASGGSSESGYPLRP 283

RESULT 7

ID 090Y29 PRELIMINARY: PRT: 310 AA.

AC 090Y29; 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE DIGEST. INTESTINAL SERINE PROTEASE.
 GN DISP OR DISP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20246299; PubMed=10786627;
 RA Shaw-Smith C.J., Coffey A.J., Leverisha M., Freeman T.C., Bentley D.R.,
 RA Walters J.R.;
 RT "Characterization of a novel murine intestinal serine protease,
 RT DISP.";
 RT Biochim. Biophys. Acta 1490:131-136(2000).
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
 -1- TRYPSIN FAMILY.
 EMBL: AJ243866; CAB56465.1; -
 HSSP: P00763; IDPO.
 DR MEROPS: S01.042; -
 DR MGD: MGI:1353645; DISP.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYPSIN_SPC.1.
 DR PROSITE: PS00240; TRYPSIN_DOM.1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER.1.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 310 AA; 33701 MW; F828EC7F6D25303F CRC64;

Query Match 32.8%; Score 564; DB 11; Length 310;
 Best Local Similarity 42.4%; Pred. No. 4e-44;
 Matches 117; Conservative 48; Mismatches 95; Indels 16; Gaps 9;

21 LLLASTALINARIIPV-PPACGPOQLNRVVGEDSTDEMPWIVSIQ---IKNGTHHCAG 77
 10 LLLTLL---QILTRAGDILPSVCGSHKADKIVGGDADLEGQWPMOVSIWTTEDG-HICGG 65

78 SLTSSRWITTAHCFKDNKRYLFSVLLGAMOLGNPGRSKQVAVWVERHPVYSKMG 137
 66 SLIHEVAVLTAAHCFKRSNL-PSFYHVKVGLTSLLEPHSTLVAVRNILFVHPTYLMA 124
 138 ACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWSIGMSIODGVPLPHPTLOK 197
 125 SSGDIALVOLDTPLRPSQ-PTPVCLPAAGRPPLTPGVYCWVWGATQOE---RDMASVLD 180
 198 LKVPIDSEVCSHLV-WRG---AGGPTEDMLCAGYLEGERDACLDSGGPLMCOVD 253
 181 LAVPLDSEDECKMYHTHQSLSGERITQSDMLCAGYVEBHIDSCGDSGGLVCSINS 240
 254 WLAGITISNCGCAERNRPVYISLSAHSRWKIV 289
 241 WTQVITSMGICARPYRPGVYTRPVYVMIORIL 276

RESULT 8

ID 09XSM1 PRELIMINARY: PRT: 273 AA.

AC 09XSM1; 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE TRYPSIN (EC 3.4.21.39).
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID=9940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=bone marrow;
 RX MEDLINE=20308142; PubMed=10848900;
 RA Pemberton A.D., McLeese S.M., Huntley J.F., Collie D.D.S.,
 RA Scudamore C.L., McEuen A.R., Walls A.F., Miller H.R.P.;
 RT "cDNA sequence of two sheep mast cell tryptases and the differential
 RT expression of tryptase and sheep mast cell proteinase-1 in lung,
 RT dermis and gastrointestinal tract.";
 RL Clin. Exp. Allergy 30:818-863(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL: Y18223; CAB41988.1; -
 DR HSSP: P20231; IAAO.
 DR MEROPS: S01.015; -
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin.1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYPSIN_SPC.1.
 DR PROSITE: PS00240; TRYPSIN_DOM.1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER.1.
 KW Hydrolase; Serine protease. TRYPTASE.
 FT CHAIN 29 273
 SQ SEQUENCE 273 AA; 30156 MW; 88A19DC779053898 CRC64;

Query Match 32.5%; Score 559.5; DB 6; Length 273;
 Best Local Similarity 40.9%; Pred. No. 8.8e-44;
 Matches 113; Conservative 53; Mismatches 95; Indels 15; Gaps 7;

21 LLLASTALINARIIPVPPACGPOQLNRVVGEDSTDEMPWIVSIQ---KNGTHHCAG 77
 2 LHLTLLALLTSL--VSAAPGQALQSGILGKEAPGSRPMQVSLRVADQYWRHCCG 59
 78 SLTSSRWITTAHCFKDNKRYLFSVLLGAMOLGNPGRSKQVAVWVERHPVYSKMG 137
 60 SLIHPQWVLTAAHCFGLPEPSDFRVOLREOH1---YQDRLLIPISRYIPRHYYWVEN 116
 138 ACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWSIGMSIODGVPLPHPTLOK 197
 117 G-ADIALQLLEPVYSICHVAKRPVTLIPRASETPPGSCQWVIGKGVNDGNRPPLPPLKO 175

Query Match	31.2%; Score 537; DB 11; Length 799;
-------------	--------------------------------------

```

YQ      41  CGKQQLNRVYGGEDSDTSDEPWIVISIQKNGTHHCAAGSLTTSRWITAAHCF-KDNLNKP  99
        ||      :|||  ||: |||  ||: |||  |||  |||  |||  |||  |||  ||: ||
Db      556  CGGLOGSLSSRTIVGGVSGSEGEWPMQASIQJGRIHICGALILADRWITAAHCFQEDSMASP  615
QY      100  YLFSVLLGAQWLQN--PGSRQKRVGVAAVEPHRVYVMKSGACADIALVLELSIQFSERY  157
        ||: |||  ||: ||: |||  ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
Db      616  KLMVTFEGKRRKQNSRNPGEVYSFKSLFLHP--YHEEDSHDIYDALQLDHPVYSATV  672
QY      158  LPQLCPDASITLPPNTHCWISGWSIQDGVPLPDPQTLOKLVPIIDSEVCSHLWRGAG  217
        |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  ||
Db      673  RPYCLPAPSHFFPEPGOHCHWTGGAREGAPVSN--TLQKVDVQLVPLQDLSGEAL-----  725
QY      218  OGPITEDMTCAGVLEDEBPACLGDSGCPLMC-QVDCATLLAGIISWGBGCAERNPQYVI  276
        : :|  |||||  :||: |||  |||||  ||: |||  |||  |||  |||  |||  |||  ||
Db      726  RLYVSPRMCLAGRKRRKDKACQDSSGSLVCRBPSSRWMLAGLVSMGLCGRPNPFQYVT  785
QY      277  SLSAHRSWVEKIT 289
        : :|  |||||  :||: |||  |||||  ||: |||  |||  |||  |||  |||  |||
Db      786  RVTVRVIMTIQOVL 798

RESULT  10
ID      Q29464      PRELIMINARY;      PRT;      237 AA.
AC      Q29464;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      TRYPLASE (EC 3.4.21.59) (FRAGMENT).
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LIVER CAPSULE;
RX      MEDLINE=96203914; PubMed=8620861;
RA      Pallaro M., Gambacurta A., Fiorucci L., Mignogna G., Barra D.,
RA      Ascoli F.,
        "cDNA cloning and primary structure of tryptase from bovine mast cells
        and evidence of the expression of bovine pancreatic trypsin inhibitor
        mRNA in the same cells.";
RL      Eur. J. Biochem. 237:100-105(1996).
CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC      TRYPSIN FAMILY
DR      EMBL; X94982; CAA64438.1; -.
DR      HSSP; P20231; IAAO.
DR      MEROPS; S01.143; -.
DR      InterPro; IPR001314; Chymotrypsin.
DR      Pfam; PF00089; trypsin.1
DR      PRINTS; PR00722; CHYMOTRYPSIN.
DR      SMART; SM00020; TRYP_SPC.1.
DR      PROSITE; PS50240; TRYPSIN_DOM.1.
DR      PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW      Hydrolase; Serine protease.
FT      NON_TER
FT      1
SO      SEQUENCE      237 AA; 26550 MW; EA537A75294EFFBA CRC64;

Query Match      31.2%; Score 536; DB 6; Length 237;
Best Local Similarity 42.6%; Pred. No. 1,1e-41;
Matches 103; Conservative 48; Mismatches 75; Indels 16; Gaps 7;

YQ      116  SRQKVVAAVEBHP-VYSWKEGACADIALVRLERSIQFSERVLPICLDPASITLPPNTH  174
        ||: |||  ||: |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  ||
Db      2      SQPMQVSLKRSYRHHHCGSSLIHQVWLVAAHCVGVEGHPSPFYRQVLEQHL---Y  58
YQ      116  SRQKVVAAVEBHP-VYSWKEGACADIALVRLERSIQFSERVLPICLDPASITLPPNTH  174

```

Tue Aug 13 16:27:48 2002

us-10-040-803-7.rspt

Page 6

```

D6      : : : : | | | | | | | : : : | | | | | | | | | |
59 YODQLPISRIIPHNCSYKNG--ADIALELDKLVNITSWHQPTLPESETEFPPTG 116

Q7      175 CMISGWSIODGVPLPHPQTLQKLPVLIDSEVCSHLYMRGAGGP---ITEDMLCAGY 230
        | | | | | | | | | | | | | | | | | | | | | | | | | |
D6      117 CWTTCGWNVDNGRRIRPPEPLKQYKVVESVSDRKTHSGLSIGDQWVPLVRDMLCAG- 175
        | | | | | | | | | | | | | | | | | | | | | | | | | |

Q7      231 LEGERDACLGSBGSLMCQVDMGLLGLISWBGCAERNRPVYISLSAHSRWEEKIV- 289
        : : | | | | | : : : | | | | | | | | | | : : : : :
D6      176 -DSGNFCQGSBGFLVCVNGTMLQAGVYSWBGCAKPRNPGLYTVNTSYLDMIHQYV 234
        | | | | | | | | | | | | | | | | | | | | | | | | | |

Q7      230 QG 291
        ||
D6      235 QG 236

```

Search completed: August 13, 2002, 08:51:08
 Elapsed time: 226 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:45:12 ; Search time 13.05 Seconds
(without alignments)
593.327 Million cell updates/sec

Title: US-10-040-803-7

Perfect score: 1720
Sequence: 1 MVSQAPALGGGCGTFTS.....AOGGALRAPSQSGAARS 317

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents.AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	581.5	33.8	299	4	US-08-944-483-66
2	554	32.2	276	2	US-09-016-366A-15
3	554	32.2	276	2	US-08-978-404B-21
4	538.5	31.3	274	4	US-08-978-404B-5
5	531.5	30.9	314	4	US-09-008-271A-3
6	530.5	30.8	274	2	US-09-016-366A-21
7	530.5	30.8	274	2	US-08-978-404B-16
8	530.5	30.8	275	2	US-09-016-366A-17
9	530.5	30.8	275	2	US-08-978-404B-12
10	529.5	30.8	273	2	US-09-016-366A-19
11	529.5	30.8	270	2	US-08-978-404B-14
12	519.5	30.2	273	2	US-08-978-404B-8
13	519.5	30.1	273	2	US-08-978-404B-3
14	514.5	29.9	267	2	US-09-016-366A-23
15	514.5	29.9	267	2	US-08-978-404B-18
16	511.5	29.7	245	4	US-08-944-483-69
17	511.5	29.7	245	4	US-09-079-970A-5
18	508	29.7	304	4	US-09-088-651-2
19	506.5	29.4	245	4	US-08-944-483-63
20	473.5	27.5	248	4	US-08-944-483-64
21	470.5	27.4	238	4	US-08-944-483-64
22	469.5	27.3	638	2	US-08-681-151-3
23	468.5	27.2	435	4	US-09-008-271A-6
24	466	27.1	454	4	US-09-518-046-2
25	463	26.9	416	2	US-09-000-846-2
26	462	26.9	492	4	US-09-342-749-2
27					Sequence 2, Appli

28	453	26.3	283	3	US-08-807-151-1	Sequence 1, Appli
29	453	26.3	283	4	US-09-478-957-1	Sequence 1, Appli
30	449	26.1	791	1	US-08-643-219-1	Sequence 1, Appli
31	449	26.1	791	3	US-08-851-350-1	Sequence 1, Appli
32	447	26.0	816	1	US-08-750-711-1	Sequence 1, Appli
33	446	25.9	546	6	5200340-6	Patent No. 5200340
34	446	25.9	790	1	US-08-469-486-54	Sequence 54, Appli
35	446	25.9	790	2	US-08-469-658-54	Sequence 54, Appli
36	446	25.9	791	2	US-09-131-995-1	Sequence 1, Appli
37	446	25.9	791	2	US-08-832-087B-1	Sequence 1, Appli
38	446	25.9	791	4	US-09-132-154-1	Sequence 1, Appli
39	446	25.9	810	1	US-07-854-603-2	Sequence 2, Appli
40	446	25.9	810	1	US-08-147-000B-29	Sequence 29, Appli
41	446	25.9	810	4	US-09-086-514-1	Sequence 1, Appli
42	446	25.9	810	6	5200340-8	Patent No. 5200340
43	445.5	25.9	256	2	US-09-027-337-3	Sequence 3, Appli
44	443.5	25.8	798	1	US-08-200-900A-2	Sequence 2, Appli
45	443.5	25.8	798	5	PCT-US94-00616-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-944-483-66
Sequence 66, Application US/08944483
Patent No. 6232456

GENERAL INFORMATION:

APPLICANT: COHEN, MAURICE

APPLICANT: COLPITTS, TRACEY L.

APPLICANT: FRIEDMAN, PAULA N.

APPLICANT: GRANADOS, EDWARD N.

APPLICANT: KASS, MICHAEL R.

APPLICANT: RUSSELL, JOHN C.

APPLICANT: STEWART, KENT D.

APPLICANT: STROUPE, STEVEN D.

TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES

NUMBER OF SEQUENCES: 76

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Abbott Laboratories

STREET: 100 Abbott Park Road

CITY: Abbott Park

STATE: IL

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSD for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/944,483

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Becker, Cheryl L.

REGISTRATION NUMBER: 35,441

REFERENCE/DOCKET NUMBER: 6183.US.01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 847/935-1729

TELEFAX: 847/938-2623

TELEX:

INFORMATION FOR SEQ ID NO: 66:

SEQUENCE CHARACTERISTICS:

LENGTH: 299 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: NO. 62324566
US-08-944-483-66

Query Match 33.8%; Score 581.5; DB 4; Length 299;
Best Local Similarity 45.4%; Pred. NO. 6.5e-50;
Matches 113; Conservative 42; Mismatches 89; Indels 5; Gaps 4;

QY 50 VVGEDSTDSMPWIVSIQKNGTHHCAGSLTTSRMVITAAHCFKDNLKPFLPSVLLGAW 109
DB 1 ITGSSANAGCMWQVITTYTGVAHVCGLVSEQWVLSAHCPESEHKE-AVEVKIGAH 59
QY 110 QINGRSRQKVGVAWVEPVPVYWKSGACADIALVLELSIQSESEVLPICLPDASIH 169
DB 60 QLDVSEDAKAVSTLKDILPHPSY-LOESQSGDIALQLSREPTISRYIRPLCPANASAF 118
QY 170 PPNHCHWISGWSIODGVPPLPHPTQLOKLVPIIDSEVCSHLYWRGA-GQGP--ITEDML 226
DB 119 PNGHCTVTGWHVAPSVSLTPKPLDOLLEVPILSRETGCLYNDKPREPHVQEDMW 178
DB 227 CAGYLEGERDACLGDGSGPLMCQYDGAWLGLIISWEGCAERNRPVGYISLSAHSRWE 286
DB 179 CAGYVEGGRKADQGGSGPLSCPYEGLWYLTGIYSWGDACGARNRPVGYTILASSYASWIO 238
QY 287 KIVGCVQLR 295
DB 239 SKVTELQPR 247

RESULT 2
US-09-016-366A-15
Sequence 15, Application US/09016366A
Patent No. 5955431

GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
APPLICANT: Huang, Chifu
TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
TITLE OF INVENTION: INHIBITORS
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 60/037,090
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/037,090
FILING DATE: 05-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7093
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441

TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 276 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-016-366A-15

Query Match 32.2%; Score 554; DB 2; Length 276;
Best Local Similarity 41.6%; Pred. NO. 3.2e-47;
Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;

QY 21 LLLASTALINARIPVP-PACGKPOQLNRVVGSDSDSEMPWIVSIQ--KNGTHICA 76
DB 6 LLLMALSLASIVYAPRPA---NQRVGIYGHASRSKPMQVSLRFLKNIWIHHC 61
QY 77 GSLTSRMVITAAHCFKDNLPFLPSVLLGAMQL--GNPGRSQKVGVAWVEPVPVYSW 134
DB 62 GSLIHQWVITAAHCVGPHIKSPOLFRLVQLRBOYLTYGD---QLLSNRTIVHPHYT 116
QY 135 KEACADIALVREISIQSERVLPICLPDASIHLPNNHCHWISGWSIODGVPPLPHPT 194
DB 117 AEQG-ADVALLLEVPVNVSTHHPISLPPASSETPPPGISCVTGTGIDNDPELP 175
QY 195 LOKLKVPIIDSEVCSHLYWRGAGQ--PTED-MICAGYLEGERDACLGDGSGPLMCQY 250
DB 176 LKQVNPVYNSLCDRKHYHTGLYTGDFFIVHGMCLAG--NTRRDSGCGDSGGLYCKV 233
QY 251 DGAMLAGITISWEGCAERNRPVGYISLSAHSRWEKIV 289
DB 234 KGTWLAGVSVWEGCAQPNKPGIYTRVYIYLDWIRIV 272

RESULT 3
US-08-978-404B-21
Sequence 21, Application US/08978404B
Patent No. 5968782

GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
APPLICANT: Huang, Chifu
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
TITLE OF INVENTION: FIBRINOGEN
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441

TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 276 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 5968782e
US-08-978-404B-21

Query Match 32.2%; Score 554; DB 2; Length 276;

Best Local Similarity 41.6%; Pred. No. 3.2e-47; Mismatches 116; Conservative 46; Mismatches 93; Indels 22; Gaps 9;

QY 21 LLLASTALNARIPVP-PACGRPOOLNRVGGEDSTDEMPWIVSIQ--KNGTHCA 76
Db LLLALSLASLVYSAPRPA---NORVGLVGHSESKMPQVSLRFRNLWNHIFCG 61
QY 77 GSLTSRVITAAHCFKNDLNKPYLFVSLGAMQL--GNPGRSQKYGVAMVEHPYYSW 134
Db 62 GSLHPQWVLAHACVGHKSPQLFRVQLREQLYLYGD-----QLSLNKLIVVHPHYXT 116
QY 135 KEGACADIALVRLERSIOFSERVLPICLPDASIHLPPTHCMISGMSIODGVPPLPQT 194
Db 117 AEGG-ADVALLLEVPVAVSTHHPISLPASFFPFGTSCWVTGMDIDNDEPLPPYP 175
QY 195 LQKLVPIIDSEVCSHLYWRGAGG--PTED-MCAGYLEGRDACLDSGGPLMCOV 250
Db 176 LKQVVPVIVENSICDRKYHTGLYTGDDFPYVHDMCLAG--NTRRDSQCGSDGGPLVCKV 233
251 DGAMLAGIISMGCAERNRPGYISLSAHRVVEKIV 289
Db 234 KGTWLDAGVSWGSCAOPNKPGLYTVTYLDMIHRY 272

RESULT 4

US-08-978-404B-5
Sequence 5, Application US/08978404B

Patent No. 5968782

GENERAL INFORMATION:

APPLICANT: Stevens, Richard L.

TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES

TITLE OF INVENTION: FIBRINOGEN

NUMBER OF SEQUENCES: 74

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: U.S.A.

ZIP: 02210-2211

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/978,404B

FILING DATE: 25-NOV-97

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/032,354

FILING DATE: 04-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: B0801/7090

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

TELEX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 274 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 5968782e

US-08-978-404B-5

Query Match 31.3%; Score 538.5; DB 2; Length 274;

Best Local Similarity 40.6%; Pred. No. 1.1e-45; Mismatches 117; Conservative 46; Mismatches 84; Indels 41; Gaps 10;

QY 21 LLLAST---AIIAARIPVPACGRPOOLNRVGGEDSTDEMPWIVSIQKNGT---HH 74
Db LLLALSLASLVYHAAHPCPKYKRVG-----IVGGRASESKMPQVSLRFRFSEMHF 57
QY 75 CASLSLSRVITAAHCFKNDLNKPYLFVSLGAMQLGNPGRSQ-----KGVAM 125
Db 58 CGSLHPQWVLAHACVGHKSPQLFRVQLREQLYLYGD-----REQLYYADOLLTVNRTV 107
QY 126 VEEHPYVSWKEGACADIALVRLERSIOFSERVLPICLPDASIHLPPTHCMISGMSIOD 185
Db 108 VHPH-YTVEDG--ADVALLLEVPVAVSTHHPISLPASFFPFGTSCWVTGMDID 164
QY 186 GVPPLHPQTLQKLVPIIDSEVCSHLYWRGAGG--PTED-MCAGYLEGRDACLSD 241
Db 165 DEPLPPYPPLKQVVPVIVENSICDRKYHTGLYTGDDFPYVHDMCLAG--NTRRDSQCGD 222
QY 242 SGGPLMCOVDMGAMLAGIISMGCAERNRPGYISLSAHRVVEKIV 289
Db 223 SGGPLVCKVAGTYLQAGVSWGSCAOPNKPGLYTVTYLDMIHRY 270

RESULT 5

US-09-008-271A-3
Sequence 3, Application US/09008271A

Patent No. 6203979

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

Hillman, Jennifer L.

Yue, Henry

Guegler, Karl J.

Cortley, Neil C.

Tang, Tom Y.

Shah, Purvi

TITLE OF INVENTION: HUMAN PROTEASE MOLECULES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/008,271A

FILING DATE: 16-Jan-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: <Unknown>

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mohan-Peterson, Sheila

REGISTRATION NUMBER: 41,201

REFERENCE/DOCKET NUMBER: PF-0458 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 314 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: PROSTUT03

CLONE: 789927

SEQUENCE DESCRIPTION: SEQ ID NO: 3 :

; TYPE: amino acid

MOLECULE TYPE: NO. 5968

MOLECULE TYPE: NO. 5968

Page 6

Matches	113;	Conservative	52;	Mismatches	86;	Indels	37;	Gaps	97
---------	------	--------------	-----	------------	-----	--------	-----	------	----

```

1  NILLALPVL--ASRAYAAPAGCALQVRGIVGGGEAPRSKWPQVSLRVHNGPYMMHECG 58

```

59 GSLHPQWLTAAHCVGPDKDLALRVQLREQHLY-----YDQQLPVRKRI 18

186 GVPLEPHPOLLOKLVPIIDSEVCSHLWRGAGQSP---ITEDMLCAGYLEGERDACLGD 24

```

242 SGGILMCQVDGAWLLAGITSWGEGCAERNRPGVYISLSAHRSWVEKIV 289
|||||:::|||||:|||||:|||||:|:::|:|:

```

Search completed. August 13, 2002, 08:47:36

Tue Aug 13 16:27:46 2002

us-10-040-803-7.ra1

